Prediction of product success: explaining song popularity by audio features from Spotify data

Author: Rutger Nijkamp
University of Twente
P.O. Box 217, 7500AE Enschede
the Netherlands

ABSTRACT

This research investigates the relationship between song data - audio features from the Spotify database (e.g. key and tempo) - and song popularity measured by the number of streams a song has on Spotify. Previous research on the topic of new product success prediction have identified multiple approaches to asking this question. Especially for products in cultural markets like music, prediction modelling is very complex. A relatively novel approach, the attribute-approach was used to explore whether song attributes have an explanatory power on stream count. Research in this specific field called Hit Song Science (HSS) has not before measured similar audio features with song popularity in stream count, which is very important for record companies and which makes this research unique. Furthermore, beneficial implications from HSS can be far-reaching to consumers, record companies and for Spotify in new value creation.

From the Spotify database API, a 1000 songs were analyzed from 10 different genres. By regression, a prediction model was built. We can conclude that our results suggest that audio features from Spotify have little to moderate explanatory power for a higher stream count, with this research design. Some significant relationships however were found, which lays a promising foundation for the research in prediction with these variables. This research contributes to further understanding in the field of HSS and the new product success prediction. Creating effective prediction models is an interesting next step to this research and so would be to expand on the variables used. Practical implications include that Spotify can further develop its database and calculations of the variables for in the future their databases will play an important role in new value creation.

Graduation Committee members:

Dr M. De Visser
Dr M.L. Ehrenhard

Keywords
Product Success Prediction Music Attributes Popularity Spotify HSS

11th IBA Bachelor Thesis Conference, July 10th, 2018, Enschede, The Netherlands.

1. INTRODUCTION

Business would be easy if we could predict product sales before they are released on the market. As the cost of failure in new product development is very high, researchers and product developers are looking for good product success/failure prediction models. Research that aims to answer these questions has established many approaches to creating success prediction models. Determinants for success are specified from organizational and industry factors, social data and predictions from tests markets to give examples. Predicting the popularity of electronic household products however seems a lot more straight-forward than predicting cultural products such as music. Their success and popularity seems related to taste and more subjective measurements which makes prediction all the more complex.

“Historically, neither the creators nor the distributors of cultural products have used analytics - data, statistics, predictive modeling - to determine the likely success of their offerings. Instead, companies relied on the brilliance of tastemakers to predict and shape what people would buy” (Davenport et al., 2009). While tastemakers are still significant influencers of products also in cultural product markets, the way how we consume is changing by technological developments. Also what we consume is changed by the shifting importance of science in art. Unprecedented access and technological advancements make prediction and recommendation of customer taste easier and more important. A specific cultural market where prediction and recommendation capabilities for producers, distributors and consumers are extremely important is the music industry.

Music is one of the most popular types of online information (Casey et al. 2008) and the importance of the music industry can be expressed in its total revenues. For 2017 they were US$17.3 billion, but it is an industry with a future too, as it is a growing industry; revenues increased by 8.1% in 2017 (IFPI, 2017). Similar to other cultural products such as movies, costs in new product development are high. Record companies are estimated to annually invest $4.5 billion worldwide in artists and repertoire (A&R) combined with marketing.

Technological advancements have seen the rise of streaming services, which revenues grew 41.1% in 2017, making digital revenues now account for more than half (54%) of the global recorded music market (IFPI, 2017). Streaming services such as Spotify are, even though the controversy on its profitability, having a positive effect on the growing music industry’s revenue (Wlömert et al., 2016) – illustrating its growing importance in the future of the music industry.

Music streaming services thank its growth to that they are able to react to new expectations of listeners, who want searchable music collections, automatic playlist suggestions, music recognition systems and more. (Casey et al. 2008). They can do so because of the (user generated) big data and their digital song database. Because this is an important value proposition for them, there are much investments made to improve it. In 2014, Spotify acquired ‘music intelligence’ company ‘The Echo Nest’ for €49.7m to further develop its service offerings like recommendation systems [1]. The music database that formed offers easily manageable data and contains all sorts of data on songs such as audio features (Tempo, Key) and track information (Artist, Genre).

Next to the importance of recommendation, there is the importance of prediction of music popularity. In the music industry too, all parties have an interest in connecting consumers with content they will like and buy and it remains one of the biggest mysteries in the industry why some songs become popular while other songs fail to do so. Researchers have started to ask the same question and many approaches to product success prediction have been taken since to predict song popularity, as will be reviewed in the literature section later.

A relatively new approach to success prediction focuses on the attributes of a product and the Spotify database, with free online available audio features, allows for this approach to be taken. With the increasing availability of digital music, the evolution of technology and the ability to retrieve information from music, a new field of research has emerged: Music Information Retrieval (MIR). MIR is a multidisciplinary domain concerned with retrieving and analysing multifaceted information from large music databases (Downie, 2003). Success prediction in this novel research field has been coined Hit Song Science (HSS), which is, as defined by Pachet (2012), “an emerging field of investigation that aims at predicting the success of songs before they are released on the market”.

This research and other research in the field have a practical dimension. The insights gained in this field can provide huge benefits for the industry and all parties involved in the music content life-cycle. Beneficial examples include that artists can work reversely the process of HSS and focus on characteristics that make their songs more popular and that record companies, aiming at maximum profit, could benefit by selecting the most promising works for publication and marketing goals (Karydis et al., 2018). Moreover, music streaming services are struggling to diversify their revenue channels and innovate on their value proposition, as can be seen by Spotify’s recent IPO. However, it is their rich databases that open up possibilities for new product success prediction models, which can create new value propositions. Examples would include they can sell prediction models to record companies and artists, but also to use them to improve their own services to music consumers. It illustrates the importance of product success prediction and the emerging research field of MIR and HSS in the field of business.

2. LITERATURE BACKGROUND

This section will further explain the relevance of this research in explaining and predicting song popularity by providing an insight in the existing body of research in product success production and specifically for music as a product in a cultural market. A gap in the research will be identified and the research question of this study will be defined.

2.1 Literature Background

Research in product success prediction models with the use of organizational- (Lo et al., 2000), industry- (De Vasconcellos et al., 1989) and entrepreneurial factors (Kleinknecht et al., 2012) are exemplary approaches to the subject. In the research mentioned, success determinants would be identified from experiences of developers, expert panels, survey data and the like. The prediction of success/failure of a new vacuum cleaner however is very different from that of a new Kendrick Lamar

[1] https://www.musicbusinessworldwide.com/spotify-acquired-echo-nest-just-e50m/

album for example, in that the content (music) does not innovate as radically on the characteristics and attributes as a new product could. Products in cultural markets, like music, ask for different determinants. There are a variety of relevant approaches that already exist in new product success prediction that are relevant for music too. Moreover, the existing body of research which defines many popularity prediction models stresses the complexity of the mechanisms of song popularity.

Let's take a look at the variety of relevant approaches that already exist. A determinant for success that can be identified by simple technology is the correlation with other items or customers. It is used in cultural markets - Spotify, as well as Netflix and Amazon are known to make use of this technology to recommend content to consumers. Naturally this method is rather limited, due to the intentions of the costumer and the nature of the shifting cultural product market over time. A weakness of this approach to work for success prediction is also that a substantial amount of customer data is needed for it to work effectively (Davenport et al., 2009).

Another approach that is implemented by companies and used in research for recommendation and prediction is the use of social networks and social data. In HSS, approaches that look for social popularity metrics include research on using social media data, for example from Twitter. Zangerla et al. (2016) found that using Twitter posts is useful to predict future charts, when recent music charts are available. Similarly, the research by Kim et al. (2014) shows a high correlation between users' music listening behaviour data from Twitter and music popularity on the charts. Furthermore, Bischoff et al. (2009) propose a music popularity prediction model by social interaction data from Last.fm, showing promising results. A weakness of this approach too is that it requires a large amount of data.

Some popular approaches in product success prediction like the use of prediction markets (Matzler et al., 2013) are not as relevant for new content as it is for products. The use of prediction markets is beneficial to take away distribution costs that do not exist for musical content (to the same extend).

Above all other approaches, the approach that has seen most popularity is found in the majority of the research in the field of Hit Song Science (HSS). It is the approach used in Music Information Retrieval (MIR) which focuses on song data - the attributes (audio features) of a song. Technological developments and user generated big data by streaming services have made this approach possible. "The underlying assumption behind HSS is that popular songs are similar with respect to a set of features that make them appealing to a majority of people. These features could then be exploited by learning machines in order to predict whether a song will rise to a high position in the chart" (Ni et al., 2015). It is important to note that all approaches described above have a weakness. The attributes approach (of MIR) requires that attributes are classified and that a lot of data is needed. An enormous amount of song data including many attributes however is already available on Spotify's database which shall be used in this research.

One of the earliest relevant work in the MIR and HSS field has been done by Dhanaraj and Logan (2005). In their study they extracted both acoustic and lyric in- formation from songs to separate hits from non-hits using standard classifiers, specifically Support Vector Machines and boosting classifiers. Their research showed promising results and found that for the features used, lyric-based features are slightly more effective than audio-based features at distinguishing hits. What is maybe surprisingly, is that they found the absence rather than the presence of certain semantic information in the lyrics mean a song is more likely to be a hit. Other work followed in 2008, as Pachet and Roy addressed a similar question regarding the prediction of popularity by automated labelling of low, medium or high popularity. They published their research 'Hit song science is not yet a science', in the, at that time, still novel field of MIR, aiming at "validating the hypothesis that the popularity of music titles can be predicted from global acoustic or human features" (Pachet & Roy, 2008). Their research found, as you might have guessed, that their learning machines weren't able to label popularity as low medium or high from audio feature sets better than random.

10 years later a broad body of research has taken the controversy on its feasibility away. Research from Lee et al. (2015) shows that it is feasible to predict the popularity metrics of a song significantly better than random chance based on its audio signal. Additionally, Ni et al. (2015) also showed that certain audio features such as loudness, duration and harmonic simplicity correlate with the evolution of musical trends. Singhi and Brown (2015) propose features from both songs' lyrics and audio content for prediction of hits and also have done research on a hit detection model based solely on lyrics' features (2014).

The growing data on listener behaviour, song audio features and meta-information provided by digital music and specifically streaming services is of great importance to MIR. Recent research is able to use database from Spotify to access real musical content easily and legally. An exemplary HSS research that uses this database is that of Herremans et al. (2014). The research focusses on the dance hit song classification problem. From a database of dance hit songs, including basic musical features as well as more advanced temporal features (timbre), classifiers were built to create dance hit prediction models. Her results suggest the possibility to predict whether a song is a 'top 10' dance hit versus a lower listed position – thus proving the capabilities of prediction models from audio features.

In the field of HSS, popularity prediction is often done in the form of hit prediction and the prediction of chart rankings. As chart rankings are not directly related to actual popularity, other measurements for popularity should be looked for.

In an attempt to predict the popularity of a song from Spotify's song data, the research of Will Berger (2017) uses (Echo-Nest) audio features similar to this research and uses Spotify's own calculated metric "popularity" to measure popularity. This is a given audio feature on Spotify's database that is computed secretly to describe the popularity of a song, while the number of streams is not given. The 'problem' with this variable however is, is that a song can score very high on popularity with only 50000 streams and vice versa – Spotify's popularity metric it does not relate to the actual number of streams. It is in this identified gap of research, and specifically with a business angle, that this research is unique.

2.2 Research Question

This paper will take the attributes approach that is data driven to test determinants for song popularity. It will address the identified gap in the existing product success prediction field of HSS by analysing stream count on Spotify instead of Spotify's popularity metric, defining popularity as hits or non-hits or by chart position. Prediction of songs in a particular genre is likely to be 'easier', since each genre has its own popular characteristics. This research wants to see whether there are general attributes for song stream count, therefore it will use songs from the 10 most popular genres as identified by Spotify. To the author's knowledge, surprisingly no research has measured popularity in the amount of streams a song has – which is of the greatest importance for record companies and artists. Here is why: Spotify pays loyalties to record companies as a fixed amount per stream on Spotify which have seen streaming revenues grown to account for 38.4% of total recorded music revenue as of 2018 (IFPI, 2018). Moreover, as streaming companies and especially Spotify are likely to become one of the biggest players in the music industry, making research into what they can do with their data also seems more relevant than ever.

Important to note is the distinction between in-sample explanatory power and out-of sample predictive power (Shmueli, 2010). This study aims to explore the in-sample explanatory power of the data already available to Spotify and developers. It will question whether the attributes approach is also found to be effective in explaining stream count on Spotify. From correlation and regression analysis, we can find if the Spotify's audio features can explain the stream count. It will be a starting point in the research for prediction of song popularity measured in stream count (out-of sample predictive power). Therefore, the research question is:

"Is the attribute approach based on Spotify's audio features effective in explaining streaming popularity on Spotify?''

3. VARIABLES AND HYPOTHESIS

3.1 Defining the Variables

3.1.1 The Dependent Variable

Since we are not discriminating between hit and non-hit song but on popularity, the dataset will have to include songs with a big variation in stream count. The popularity of a song can be measured a posteriori according to statistics such as the number of times a track has been played – in this context: streamed. We define popularity as the number of streams on Spotify. The variable number of streams is easily measurable in the Spotify application, but it has to be done by hand. The number of streams is a continuous ratio variable put in the dataset as 'Streams'.

3.1.2 The Independent Variables

Since this study is interested in what Spotify can do with their current data from a business angle, song popularity will be operationalized by attributes given by the Spotify database. The independent variables in this study consist of just of all the basic features extracted by the Spotify Search API from the online database of Spotify[2]. All the audio features available where selected, except for 'Time Signature' which is very likely to be 4/4 for every song and therefore not relevant for explanatory power (for prediction it could). Spotify keeps the formulas to calculate special features such as 'Danceability' secret (likely confidential for business reasons). Variables like 'Tempo' are fairly straight forward and variable 'Genre' was added by hand as an extra independent variable to run in the regression.

With newly available large and rich datasets come complex relationships and patterns that are hard to hypothesize, especially given theories that exclude newly measurable concepts (Shmueli, 2010). Additionally, no theories exist specifically on the attributes that are specifically calculated by Spotify secretly and other research that works with Spotify's audio features gave no hypothesis or theories on them either. For this reason, hypothesis will be derived from simple argumentations that are not tested by theory. Our correlation analysis however will give us insight as into our relationships and whether our hypothesis hold some truth.

3.1.2.1 Acousticness

Acousticness is an attribute that is calculated by Spotify and is a confidence measure from 0.0 to 1.0 representing if the track is acoustic. 1.0 represents high confidence the track is acoustic. Looking at the high number of non-acoustic popular songs, and the array of electronic instruments used in the charts, a negative relationship between Acousticness and stream count is most likely. Hence the following is hypothesized:

H1: Acousticness is negatively related to a higher stream count.

3.1.2.2 Danceability

Danceability is calculated by Spotify and describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. Regularity and whether a track is danceable is likely a characteristic of popularity.

H2: Danceability is positively related to a higher stream count.

3.1.2.3 Duration

The duration of the track measured in milliseconds. Usually popular songs are not too lengthy as they might bore the listener. Furthermore, the mean of duration in our dataset is 226 seconds and therefore we expect a higher duration to be negatively related to stream count.

H3: Duration is negatively related to a higher stream count.

3.1.2.4 Energy

Energy is a measure from 0.0 to 1.0 calculated by Spotify, to represent intensity and activity. The measure is based dynamic range, perceived loudness, timbre, onset rate, and general entropy to represent a perceptual measure of intensity and activity. Spotify provides the example that death metal has high energy while a Bach prelude scores low on the scale. Since intensity and activity are characteristics of a song that grabs the attention, it is likely that popular songs score reasonably high on energy.

H4: Energy is positively related to a higher stream count.

3.1.2.5 *Instrumentalness*

Instrumentalness is calculated by Spotify and predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the Instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Looking at popular songs, we see that most songs contain lyrics and we assume that the ability to sing along with a track can make the track more popular.

H5: Instrumentalness is negatively related to a higher stream count.

3.1.2.6 Key

This measure represents the key the track is in, represented as an integer. Integers map to pitches using standard Pitch Class notation . E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. Spotify data analyst and jazz pianist Kenny Ning explains that G, C and E are convenient keys for guitar and piano[3]. Although it is hard to make

[2] https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/

[3] https://insights.spotify.com/it/2015/05/06/most-popular-keys-on-spotify/

a hypothesis on Key, it is likely that these keys are most used and will be most often seen in popular songs.

H6: Songs in the Key of C, G and E are likely to be positively related to a higher stream count.

3.1.2.7 Liveness
Liveness is a measure from 0.0 to 1.0 that detects the presence of an audience in the recording. It is calculated by Spotify and a value above 0.8 provides strong likelihood that the track is live. By looking at popular tracks, we often see very polished studio productions. Applause and other 'live noise' would disturb the actual song and its production.

H7: Liveness is negatively related to a higher stream count.

3.1.2.8 Loudness
Loudness measures the overall loudness of a track in decibels (dB). Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typical range between -60 and 0 db. We assume that the louder the song, the better it is able to communicate emotions.

H8: Loudness is positively related to a higher stream count.

3.1.2.9 Mode
Mode indicates the modality, major (1) or minor (0) of a track, the type of scale from which its melodic content is derived. Since a Major mode sounds more cheerful than a Minor mode, it is likely that a Major mode is more popularly used in hit songs.

H9: The Major mode is positively related to a higher stream count.

3.1.2.10 Speechiness
Speechiness detects the presence of spoken words in a track, measured from 0.0 to 1.0. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0. Current popular songs have lyrics but consist of music primarily. Moderate Speechiness – songs that contains both music and lyrics - is likely positively related to a higher stream count. A high Speechiness is not, and thus it is hypothesized that:

H10: Speechiness is negatively related to a higher stream count.

3.1.2.11 Tempo
Tempo measures the overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration. Given that rap and hip-hop, which tempo is quite low, are part of main stream chart music, it is difficult to say whether tempo currently has much meaning for predicting song popularity.

H11: Tempo is not significantly positively/negatively related to stream count.

3.1.2.12 Valence
Valence is a measure from 0.0 to 1.1 calculated by Spotify describing the musical positivity conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry). By looking at current popular songs we can assume that popular tracks are generally more cheerful as the most likely reason why a person would listen to popular music is to cheer him/her up.

H12: Valence is positively related to a higher stream count.

4. METHODOLOGY

In this chapter, the methodology of the research is explained. This will be done in the following order: defining the research population and sample, elaborating on the data collection method and data pre-processing.

This research will use and analyse data retrieved by Spotify API database as of 2018. Spotify provides this Application Programming Interface (API) as an open and free database for developers to use and build their applications. Spotify's Search API allows you to obtain audio features for any song on Spotify based on a search, where a number of parameters can be selected, including genre and year.

4.1 Defining the Population and Sample

Songs on Spotify are our unit of observation. These songs need to be drawn from a sample from a certain population to be relevant to our research question.

Our sample needs to be drawn from a relevant population and in order to construct this; the dataset will include any popular genre, which can be defined as music with a wide appeal. Therefore, the 10 most popular genres on Spotify are selected, which are, as specified by Spotify: "Pop", "Rap", "Dance-pop", "Hip-hop", "Rap-pop", "Post-Teen-Pop", "Rock", "Modern-Rock", "Trap Music" and "Latin". Each song on Spotify is labelled with a genre, although a song can have multiple genres.
Furthermore, a time frame of released songs needs to be selected. By specifying the year parameter as 2017, we could gather a wide range of tracks from the past year. 2017 was decided as a year for two reasons. One reason is that selecting a certain year takes away a bias towards songs that are released earlier, since they naturally have had more time to solicit streams. Another reason is that 2017 is the most relevant year to give us information about what currently is popular in music as the dominant music that people listen to changes over time (Herremans, 2014).

To get a representative sample of popular and non-popular tracks across the selected genres, 60 songs were pulled from each genre with Spotify's Search API. The top 20 most "popular" (by Spotify's own metric) were selected and 250 songs were skipped using offset to select another 20 four times. This method gives a range of popularity both across genres and within genres. Since we selected 10 genres, 1000 songs were pulled.

4.2 Collecting the Data

For collection, the Spotify database was used. The free database is an online API that is open for developers to use data and build their own applications. Spotify's Search API allows you to obtain audio features for any song on Spotify based on a search, where a number of parameters can be selected, including genre and year. A python script was created to automatically pull the data from the database – the Spotify search can only request data one song at a time – and create a dataframe. Spotipy[4], created by Paul Lamere (2014), is a lightweight Python library for the Spotify Web API. It was used to write the script to query songs from the Spotify database. Each song's audio features were pulled with the "get audio features" API request. These features will make up the independent variables to research.

Next, Python corralled all the data into a pandas dataframe, which is a matrix with labels that can be opened in Excel. Each row contains a track, while each column contains the values for

[4] http://spotipy.readthedocs.io/en/latest/#

the audio features. The number of streams was tracked by hand for each song in the Spotify application as it is not a feature that is given in the API. Since a song can have multiple genres in Spotify's database, there were duplicates to remove. After removing 91 duplicates, 901 songs would make the database.

4.3 Data Pre-Processing

The following steps are explained to gain insights into the data and for data pre-processing.

4.3.1 Cleaning the data

Each song feature is not measured equally in the Spotify API. Predictors 'Acousticness', 'Danceability', 'Energy', 'Instrumentalness', 'Liveness', 'Speechiness' and 'Valence' are all measured from 0-1 and are (continuous) ratio variables. 'Liveness' was recoded in a dummy variable where only the cases above 0,8 where selected as 1 (As suggested by Spotify), with all other cases as 0. Variables 'Duration' (Ms), 'Loudness' (Db) and 'Tempo' (BPM) are also ratio variables measured in their own units. Secondly, our predictor 'Key' is measured as a (categorical) nominal variable with 0-11 corresponding to the keys according to the standard Pitch Class notation (0=C, 1=C#). Similarly measured is 'Genre', which is also a nominal variable (1=pop, 2=rap). A third distinct variable is 'Mode' which is a dichotomous variable, where 0 is defined as the Minor mode and 1 as Major. In order to be able to measure the correlation of the separate categories in our nominal variables 'Key' and 'Genre', they were recoded into separate binary dummy variables in SPSS and added to the regression after the pandas dataframe was imported in SPSS.

4.3.2 Assumptions for Regression

In SPSS, a linear regression will be performed. Linear regression is an analysis that assesses whether the predictor variables, the audio features, explain the dependent variable which is the number of streams. A linear regression is a parametric statistical test that can only be done when five key assumptions are met: linear relationship, multivariate normality, no (or little) multicollinearity, no auto-correlation and homoscedasticity. The distribution of the dependent variable plays was plotted (Normal plot, Q-Q plot and Residual vs. Predicted value plot) and looked much skewed to the left. Because the variable is not normally distributed it does not meet the assumptions of parametric statistical. Furthermore, heteroscedasticity can be observed from the plots. To make the dependent variable more normal and eliminate heteroscedasticity, a data transformation was performed. Data transformation essentially entails the application of a mathematical function to change the measurement scale of a variable that optimizes the linear correlation between the data. The function is applied to each point in a data set — that is, each data point y_i is replaced with the transformed value. The dependent variable 'streams' is a continuous variable where there is much variation in the counts of the most popular songs and the least popular songs. To address for this relative change, a log transformation is performed on the variable 'streams' to pull the data closer together. Gelman and Hill (2007) wrote that natural logs (logarithms base ***e,*** abbreviated as ln) are preferred for coefficients on the natural-log scale are directly interpretable as approximate proportional differences.

The new distribution (y'=ln(y)), the variable 'StreamsLn' is normal enough to continue with the linear regression. The transformed variable also now meets the assumption of homoscedasticity.

5. ANALYSIS

In this section we aim at stating the findings of the analysis using several measures. First, by correlation analysis, hypothesis are tested. Afterwards, by linear regression analysis, variables are selected for a prediction model to measure the explanatory power of our audio features for stream count.

5.1.1 Correlation Analysis

In Table 1 the Pearson correlations of the audio features and the dependent variable 'Streams' are presented and the significance (2 tailed) below. For convenience of viewing, only the significant keys of the variable Key were selected. Now we see if the relationships in the hypothesis have found to be significant by a correlation analysis.

H1: Acousticness is negatively related to a higher stream count.

With a significant r = -0,128, there is evidence to accept the hypothesis.

H2: Danceability is positively related to a higher stream count.

With a significant r = 0,231, there is evidence to accept the hypothesis.

H3: Duration is negatively related to a higher stream count.

With a significant r = -0,155, there is evidence to accept the hypothesis

H4: Energy is positively related to a higher stream count.

With a non-significant r = -0,015, there is no evidence to accept our hypothesis and the relation is found to be negative instead.

H5: **Instrumentalness is negatively related to a higher stream count.**

With a significant r = -0,156, there is evidence to accept the hypothesis

H6: Songs in the Key of C, G and E are likely to be positively related to a higher stream count.

There were no significant relations found for the key of C, G and E. Instead for the key of D a significant r = -0,064 and for B a significant r = 0,105 was found. We are unable to accept our hypothesis.

H7: Liveness is negatively related to a higher stream count.

With a significant r = -0,167, there is evidence to accept the hypothesis.

H8: Loudness is positively related to a higher stream count.

With a non-significant r = 0,064, we are unable to accept the hypothesis.

H9: The Major mode is positively related to a higher stream count.

With a non-significant r = 0,039, we are unable to accept the hypothesis and the relation is found to be negative instead.

Table 1. **Correlations**

	ACOU	DANC	DUR	ENER	INSTR	LIVE	LOUD	SPEECH	TEMP	VAL	D	B	MAJ	MIN
Streams	-,128	,231	-,155	-,015	-,156	-,167	,064	,068	,034	,031	-,074	,105	-,039	,039
	,000	,000	,000	,643	,000	,000	,055	,041	,312	,354	,027	,002	,247	,247

H10: Speechiness is negatively related to a higher stream count.

With a significant r = 0,068, we must reject the hypothesis and conclude that the relationship is found to be positive.

H11: Tempo is not related to stream count.

With a non-significant r = 0,034, we accept the hypothesis that tempo is not related to stream count.

H12: Valence is positively related to a higher stream count.

With a non-significant r = 0,031, there is no evidence to accept the hypothesis.

5.1.2 Regression Analysis

After testing the hypothesis for each variable, the question remains whether a combination of independent variables can lead to a reliable regression model. In this section, a linear regression model is built with SPSS to do so. To run a regression analysis, a regression model needs to be specified. By using the step-wise method, only the relevant variables with statistically significant correlations are entered and selected. The selection reduces the data to 9 attributes. Removing insignificant determinants and only selecting predictive variables improved the explanatory power, the R^2 of our model slightly. This is needed for our model to be 'true' in reality – to our data. Important to understand is that models for prediction work different. They do not have to be 'true', as long as their predictive power is higher (Schmueli, 2010). Thus the following model is by no means similar to the model with the best predictive power.

The empirical model used to examine the determinants of stream count can now be specified as:

$$StreamsLn = \beta0 + \beta1Pop_i + \beta2Danceability_i + \beta3Rock_i + \beta4Acousticness_i + \beta5Liveness_i + \beta6Rap_i + \beta7Dance\text{-}pop_i + \beta8Instrumentalness_i + \beta9KeyB_i$$

where

Pop, Rock, Rap, Dance-pop, and Rock are dummy variables indicating the genre of a song and where KeyB is a dummy variable indicating the key of a song. The variables 'Duration', 'Energy', 'Loudness', 'Speechiness', 'Mode', other 'Genre' dummies, 'Key', 'Tempo' and 'Valence' where found not to be predictive variables (by step-wise method) for the stream count 'Streams'.

Table 2. **Model Summary**

Model	R	R^2	Adjusted R^2	Std. Error Estimate
9	,458[a]	,210	,202	1,67013

h. Predictors: (Constant), Pop, Danceability, Rock, Acousticness, Liveness, Rap, Dance-pop, Instrumentalness, KeyB

B. Dependent Variable: StreamsLn

In Table 2 we see the Model summary of the regression model, where the explorative power of our regression model to the stream count ('StreamsLn') is given as Adjusted R^2 = 0,202. This means that the regression model accounts for 20,2% of variation in 'StreamLn', the stream count. After checking residual plots and the ANOVA output where we find a significant p value (<0,05) for the F test, we can conclude that the model provides a better fit than the intercept-only model.

The regression model is run in SPSS and the coefficient matrix can be seen in Table 3. Statistical conclusions can be derived from this SPSS output. Since the dependent variable is log natural transformed, the beta coefficients need to be transformed back in order to be able to interpret them correctly. The estimated coefficient of the 'Acousticness' variable is for example β1 = -0,117 so we would say that an increase of one-unit in 'Acousticness' would result in ($e^{\beta 1}$-1) x 100 = approximately

Table 3. **Regression Coefficients**

Model		Unstandardized Coefficients		Standardized Coefficients			Transformed Standardized Coefficients in percentage
		B	Std. Error	Beta	t	Sig.	
9	(Constant)	15,354	,313		49,058	,000	
	Pop	1,691	,183	,284	9,238	,000	+33%
	Danceability	1,421	,406	,116	3,498	,000	+12,3%
	Rock	-,650	,197	-,109	-3,291	,001	-10,3%
	Acousticness	-1,038	,267	-,117	-3,891	,000	-11%
	Liveness	-1,269	,366	-,106	-3,467	,001	-10,1%
	Rap	,728	,197	,114	3,694	,000	+12,1%
	Dance-pop	,582	,187	,097	3,118	,002	+10,2%
	Instrumentalness	-1,385	,524	-,081	-2,645	,008	-7,8%
	KeyB	,512	,200	,077	2,553	,011	+8%

a. Dependent Variable: StreamsLn

-11% change in Y. The transformed standardized coefficients are also added to the SPSS output found in Table 2.

'Acousticness' is a confidence measure, measured from 0-1 whether the track is acoustic, so our results tell us that if a track is acoustic, the stream count decreases by 11%. A 1% increase for more continuous variables measured between 0.00 and 1.00, are divided by 100 for correct interpretation. A 1% increase in 'Danceability' leads to a 12,3/100 = 0,123% increase in the stream count. If a track is instrumental, this will impact the stream count negatively by 7,8%. If a track is very likely to be live, this will impact the stream count by -10,1%. If a track is in the genre of Pop, the stream count will increase with 33%, for Rap by 12,1%, for Dance-pop by 10,2% and for Rock it will decrease by (-)10,3%. This is in the line with expectation that Pop music remains the most popular genre and that there is indeed a growing popularity for the genre of Rap. Contrary to our hypothesis however, it was found that the key of B is significant to increase the stream count by 8%.

6. DISCUSSION

This research intended to answer the research question *'Is the attribute approach based on Spotify's audio features effective in explaining streaming popularity on Spotify?*'. The question was analyzed by using Spotify's audio features taking an attribute based approach to a success prediction model for the number of streams a song has on Spotify. The results of the correlations showed that there were significant relationships and that the directions of the relationships seemed to fit our hypotheses. The found relationships however were generally weak. Next a regression model was built from a selection of attributes by a step-wise method. Its explanatory power (R^2) is 20,2%, meaning that the model explains 20,2% of the variation in the stream count. We can conclude that our model is not as effective in explaining stream count on its own.

There are some reasons that likely limit the explanatory power of our model. It could be mainly due to the question we are asking. Since all genres were included for measuring streaming popularity, it is likely that, because different genres do not share the same popular attributes, there will be noise in the hit prediction model making for a lower R^2 value and lower correlations. Research from Herremans (2014) has been successful in predicting whether a dance song can become a hit songs – addressing the importance of genre specification in the research. By looking at the results of this research, other research in the field of HSS and new product success prediction, we come to understand that the best model for explaining and predicting song popularity, also measured in stream count, performs a balancing act. They include 'internal' features, audio features and meta-data, and 'external' data, tastemakers and social media interactions.

6.1.1 Academic Implications

The research has primarily helped to further explore whether the attribute approach in new product success prediction with Spotify's audio features can explain our popularity measure, Spotify's stream count. No research before had used stream count as the popularity measure, which is as of 2018 of the greatest importance for record companies.

Our explanatory power of 20,2% does not say much about the predictive power of the audio features, but can better be seen as a starting point for prediction since our analysis shows promising results for prediction with audio features. Future research can develop the work by creating prediction models from these features (decision tree, support vector machines) for predicting song popularity.

Future research could also expand on this work by including other established predictors such as social media data in creating explanatory or prediction models that have better predictive power. The model can also be created to predict songs in specific styles and contexts of music.

6.1.2 Practical Implications

The found relationships and regression model can be seen as the biggest practical contribution of this research. More tests might be needed to improve the explanatory power of the model. With the academic implications that it has brought, new findings can create opportunities from where new value creation for Spotify can arise. Examples include selling hit prediction models to record companies and artists. Spotify might be interested in also collecting other 'external' data for this reason and possibly add social functionalities such as a chat box. Especially now, as streaming companies like Spotify are struggling with their profitability, diversifying the revenue channels and innovating on the value proposition is vital.

This research adds to the exciting findings of the broad body of research in the attribute approach and Hit Song Science. Success prediction in cultural markets and especially for products remains a very complex subject. The novel possibilities of the use of data will make for an interesting transition in the years to come.

7. ACKNOWLEDGMENTS

I would like to thank my supervisor Dr. Matthijs de Visser for his help and feedback with this thesis. Without his understanding and patience I could not have chosen a thesis subject that was so interesting to work on.

8. REFERENCES

Berger, W. (2017). Why is this song Popular? (Feat Spotify). Retrieved 01/06/2018 from: https://medium.com/@albert.w.berger/what-makes-a-song-popular-in-a-certain-country-feat-p%CC%B6i%CC%B6t%CC%B6b%CC%B6u%CC%B6l%CC%B6l%CC%B6-spotify-cd705abc59

Bischoff, K., Firan, C.S., Georgescu, M., Nejdl, W., & Paiu, R. (2009). Social knowledge-driven music hit prediction. Proceedings of International Conference on Advanced Data Mining and Applications, 43-54.

Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M. (2008). Content-Based Music Information Retrieval: Current Directions and Future Challenges, 668-669.

Davenport, T.H., & Harris, J.G. (2009). What people want (and how to predict it). MIT Sloan Management Review vol. 50 no. 2, 24-27.

Dhanaraj, R., & Logan, B. (2005). Automatic prediction of hit songs. Proceedings of Conference on International Society for Music Information Retrieval, 488–491.

Downie, J. (2003). Music information retrieval. Annual Review of Information Science and Technology, 37(1), 295–340.

De Vasconcellos, E.S., J.A.S., Hambrick, D.C. (1989). Key success factors: Test of a general theory in the mature industrial-product sector. Strategic Management Journal Volume 10, Issue 4, 367-382.

Gelman, A., & Hill, J. (2007). Data Analysis using Regression and Multilevel/Hierarchical Models. Cambridge University Press: Cambridge; New York, 60-61.

Herremans, D., Martens, D., & Sörensen, K. (2014). Dance Hit Song Prediction, 291-293.

IFPI. (2017). Global Music Report 2017.

IFPI. (2018). Global Music Report 2018.

Karydis, I., Gkiokas A., Katsouros V., & Illiadis L. (2018). Musical track popularity mining dataset: Extension & experimentation, 1.

Kim, Y., Suh, B., & Lee, K. (2014). The future billboard: mining music listening behaviors of Twitter users for hit song prediction. Proceedings of International Workshop on Social Media Retrieval and Analysis, 51–56.

Kleinknecht, A.,& Van der Panne, G. (2012). Predicting new product sales: The post-launch performance of 215 innovators.

Lee, J., & Lee, J.S. (2015). Predicting music popularity patterns based on musical complexity and early stage popularity. Proceedings of the Third Edition Workshop on Speech, Language and Audio in Multimedia, 3-6.

Lo, F.C.-W., Foo, S.W., & Bauly, J.A. (2000). Multiple regression models for electronic product success prediction. Proceedings of the 2000 IEEE International Conference on Management of Innovation and Technology: Volume 1, 2000, Article number 917374, 419-422.

Matzler, K., Grabher, C., Huber, J., & Füller, J.a,c. (2013). Predicting new product success with prediction markets in online communities. R and D Management Volume 43, Issue 5, 420-432.

Ni, Y., Santos-Rodriguez, R., Mcvicar, M., & De Bie, T. (2015). Hit song science once again a science?, 1-2.

Pachet, F. (2012). Hit song science.

Pachet, F., & Roy, P. (2008). Hit song science is not yet a science. In J.P. Bello, E. Chew, and D. Turnbull, editors, Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR), 355-360.

Shmueli, G. (2010). 'To explain or to predict?'. Statistical Science Vol. 25, No 3, 289-310.

Singhi, A., & Brown, D. (2015). Can song lyrics predict hits? Proceedings of the 11th International Symposium on Computer Music Multidisciplinary Research, 457–471.

Wlömert, N., & Papies, D. (2016). On-demand streaming services and music industry revenues – insights from Spotify's market entry. International Journal of Research in Marketing (33-02), 314-327.

Zangerle, E., Pichl, M., Hupfauf, B., & Specht, G. (2016). Can microblogs predict music charts? An analysis of the relationship between #nowplaying tweets and music charts.