How to produce a Hit Song in 2022

Analysis of the determinants of chart success over the past decades

Bachelor Seminar Paper by Rouven Beiner

Universität Leipzig

Wirtschaftswissenschaftliche Fakultät

Institut für Empirische Wirtschaftsforschung

Professor Bernd Süßmuth

Leipzig, 27.03.2022

Matrikelnummer: 3716786

Email-Adresse: rouven.beiner@gmail.com

Telefonnummer: 0157 32451762

Adresse: Lampestr. 9, 04107 Leipzig

Table of Contents

1. Introduction

The charts continue to be a reliable success indicator in the music industry. In addition to retail and download sales, streaming revenues have also been integrated for some years now (GfK Entertainment). A chart placement does not only reflect high previous earnings but can increase sales of current and future work (Strobl & Tucker, 2000). Moreover, it can trigger so called bandwagon effects which increase with rising chart positions. “Developed initially by Leibenstein (1950), bandwagon effects arise when the desirability of a course of action depends positively on the number of other people who are expected to undertake the same action.” (Strobl & Tucker, 2000, p.118-119). It is not without reason that artists are interested in knowing the determinants influencing the chart success of their songs.

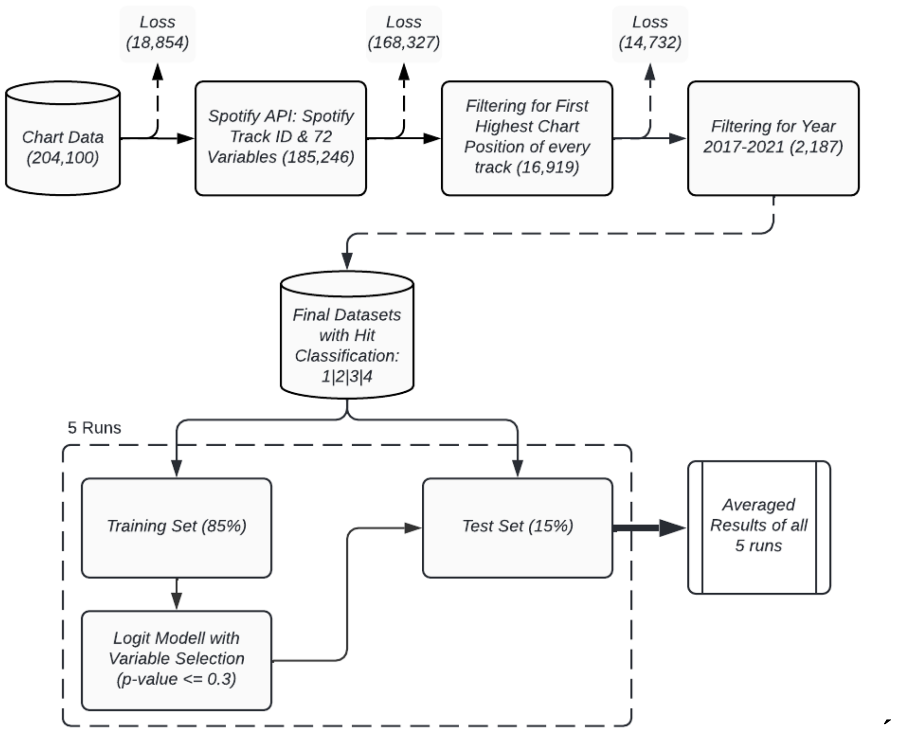
The analysis of the german charts and 72 variables shows that some of the variables have a significant influence on the chart success. Given the set of regressors, with four logistic regression models, I calculate a songs probability of reaching a top Chart Position. The best logit model can predict Hit-Songs with an accuracy of 69%, which is clearly above that of random assignment (50%). In some studies, authors were able to achieve a higher accuracy by using other algorithms such as Support Vector Machines or Random Forest (Herremans et al., 2014; Georgieva et al., 2018; Middlebrook & Sheik, 2019), additional variables, lyrics analysis (Singhi & Brown, 2014) or a different classification of Hits and Non-Hits.

2. Related Work

As early as 1988, the successful producer duo „The KLF“ published a book named "The Manual (How to Have a Number One the Easy Way)." The step-by-step guide may not be the result of empirical research, but at least it shows that the success of songs and artists is not a product of chance and can be influenced. In contrast, the Hit Song Science (HSS), which has been established in recent years, attempts to determine the factors of chart success and characteristics of Hit-Songs through empirical research. Previous studies in the field of HSS have shown that quite a few musical characteristics have a significant influence on the probability of charting (Interiano et al., 2018; Georgieva et al., 2018; Herremans et al., 2019; Middlebrook & Sheik, 2019). When analysing the german charts, I came to similar results.

3. Data

Table 1: Flowchart Data Collection



The basis of the analysis are the german weekly top 100 (from 08/1989) and top 75 (before 08/1989) charts since 1980 published by GfK Entertainment on www.offiziellecharts.de. The database contains 204,100 songs. Using the Spotify API, I was able to obtain the 72 variables. Almost all chart songs could be found in the Spotify catalogue. 9% of the songs were not found because not all tracks are available in the Spotify catalogue. The peak chart position is a good measure of a song's success. Since most songs are in the charts before and after the first chart peak with a lower position, all lower chart entries were removed. After this filtering, the database contains 16,919 unique songs, which were used for the following calculations. Table 3 shows the annual averages of selected variables. For the average song length (duration), a conspicuously strong downward trend can be seen. One explanation for this development may be the so-called "play-share" remuneration structure of music streaming services. Artists receive royalties as soon as a song is played for at least 30 seconds. This creates an incentive to shorten the song length in order to maximize streaming play counts and earnings (Glenn Mcdonald, 2021).

Table 2: Histogram Observations per year & Chart Rank

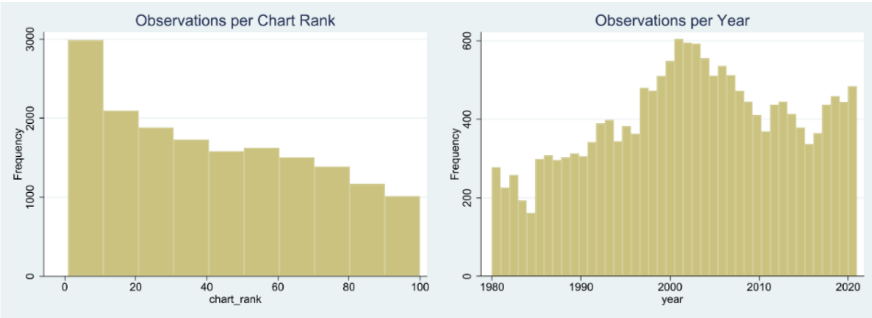


Table 3: Yearly Means of chosen variables

(Valence, Mode, Duration, Loudness, Speechiness, Danceability, Acousticness, Energy, Liveness)

4. Exogenous Variables: Socioeconomic and musical Determinants.

The determinants of chart success can basically be divided into factors of a socio-economic and musical nature. The former includes, for example, the initial popularity of the artist or the marketing budget (Strobl & Tucker, 2000). The set of musical factors can be divided into three subgroups: 1# track & artist information, 2# audio features calculated by Spotify, and 3# tonal characteristics from audio spectral analysis.

The track & artist information includes variables such as track length, volume, or genres with which the artist can be associated. I suspect that the first section in particular is important as a "first impression". Therefore, the length, volume, and mode of the first section are calculated. The variables of the second category are audio features calculated by Spotify. They take values between 0 & 1 and describe specific "tangible" features of the song (like Danceability, Valence, Speechiness). The underlying calculation algorithm has not been published; however I think that the 12 Timbre Vectors play a major influence. The third category (the 12 Timbre Vectors) are abstract parameters describing timbre. They are roughly centered around zero. To classify the timbre of a song, Spotify defined 12 function and patterns in the audio spectrogram. Each of these patterns represents a specific tonal character such as overall volume (timbre 1), brightness (timbre 2) or attack (timbre 4). The audio spectrogram is a representation of pitch, volume and time. Because of the time dimension, the 12 timbres must be calculated for each segment of a song. A segment contains a roughly consistent sound throughout its duration. To get the songs overall timbres, I extracted statistical moments (mean, standard deviation, minimum and maximum) of the 12 timbre vectors of all segments. Herremans et al. (2019) have also included these statistical moments of the timbres in their model and were able to achieve very good results. One of the socioeconomic variables is the "Superstar Variable". It assumes that songs have a higher hit probability if the artist has achieved previous chart success. In this analysis, the superstar variable is a dummy that is 1 if the artist has at least 3 previous chart entries. Previous studies “found that the prediction accuracy can be greatly improved when the superstar variable is included” (Interiano et al., 2018).

Table 3: Exogeneous Variables

Track & Artist Information (14)	**Description**
Explicit	Dummy Variable. Whether or not the track has explicit lyrics
Genre (10)	Categorical Variable. Genres the artist is associated with. (Set by Spotify)
Duration	The duration of the track in milliseconds.
Key	Categorical Variable. The key the track is in. Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. If no key was detected, the value is -1.
Mode	Dummy Variable. Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.
Tempo	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.
Loudness	The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typically range between -60 and 0 db.
Section 1 (3)	Calculation of Duration, Mode, and Loudness for the first Section of the Track

Audio Features (7)	**Description**
Danceability	Danceability describes how suitable a track is für dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.
Energy	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.
Instrumentalness	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. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.
Liveness	Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.
Speechiness	Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.
Valence	A measure from 0.0 to 1.0 describing the musical positiveness 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).
Acousticness	A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.

Audio Analysis (48)	**Description**
Timbre 1-12	Calcultaion of Statistical Moments (Mean, Standard Deviation, Minimum, Maximum) of the 12 Timbre Vectors.

Socioeconomic (3)	**Description**
Popularität („Superstar“)	Categorial Variable indentifies the number of prior chart entries of an artist.
Major	Dummy Variable identifies if the publishing label is one of the three majors (Sony, Warner, Universal)
Feature	Dummy Variable identifies if the track have a featured artist

5. Methods & Models

i. Linear Regression w. Interaction Terms / Marginal-Effects

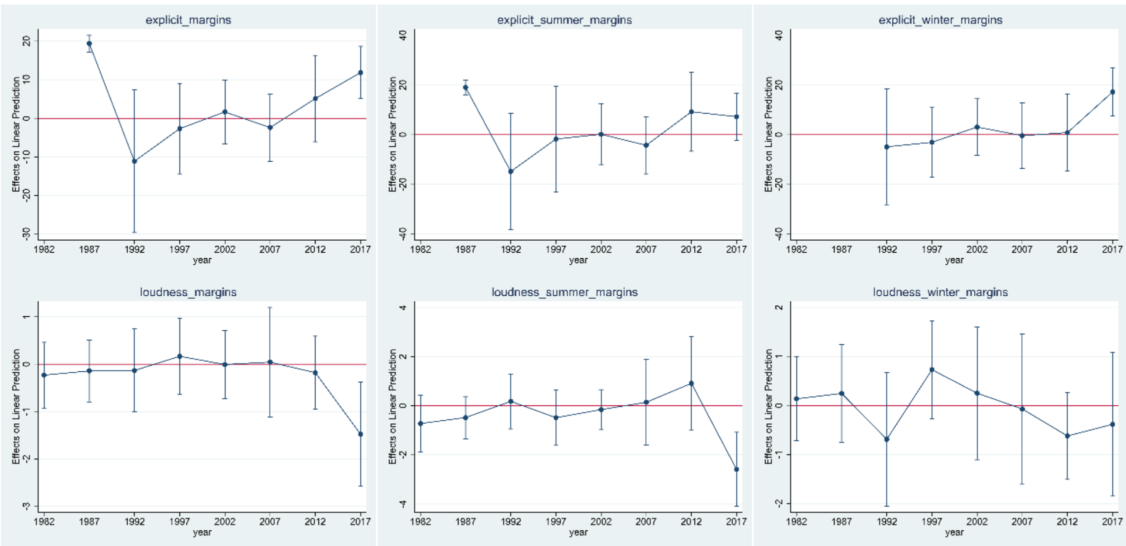
A simple linear regression model is used that includes main effects as well as interaction terms of the respective variable and year. The Chart position is the dependent variable.

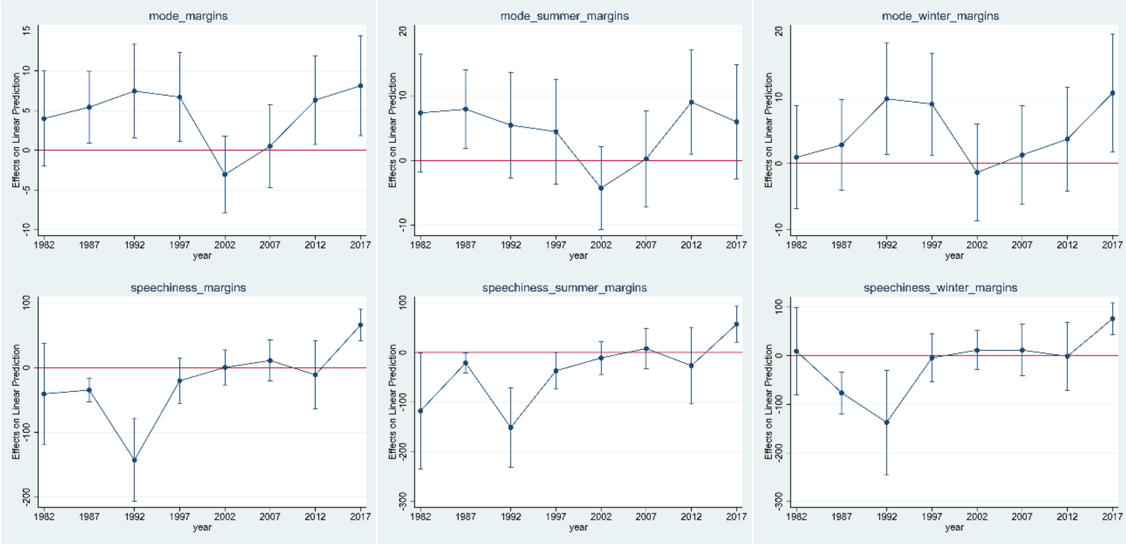
$$\mathbf{Y = \alpha + \beta_1 \times variable + \beta_2 \times year + \beta_3 \times variable * year + \varepsilon}$$

Year represents the five-year periods starting in 1982. In the figures, the beta coefficient of the regression for each five-year period is shown with the corresponding 95% confidence intervals. Values above the red zero line represent a negative impact, and values below the zero line represent a positive impact. I found that the season is an important influencing factor, so for each variable, in addition to the total effect, the winter and summer effect was calculated (summer = april to september). In the last 5-year period from 2017 to 2022, a volume increase of 1 DB improves the chart position by 1.5 positions. However, this impact is only significant in the summer months. Although the average speechiness in the songs has been increasing strongly since 2012 (Table 3), a high speechiness now no longer influences the chart position positively (as in 1992 to 1997), but negatively.

Table 5: Margins of Variables per Year

(Explicit, Loudness, Mode, Speechiness)





ii. Logit Probability Model

Table 6: Model Results

	Classification		Accuracy					Test Set		Training Set	
	Hits	Non Hits	Overall	Hits	Non Hits	delta	R2	Hits	Non Hits	Hits	Non Hits
1	Top 10	50-100	64%	27%	85%	13%	14%	72	125	359	738
2	Top 10	80-100	63%	69%	55%	**24%**	**18%**	72	48	359	294
3	Top 20	50-100	58%	48%	**68%**	16%	10%	112	125	598	738
4	Top 20	80-100	**69%**	**85%**	31%	16%	13%	112	48	598	294

For the logistic regression models, I used the filtered database shown in Table 1 for the period starting in 2017. The four models differ only in the definition of hits and non-hits. First, all variables were added, and then the most insignificant variable in each case was removed (backward reduction) until only variables with a P-value of <=0.3 were included. The resulting set of relevant variables is different for each model due to the diversity of the underlying training sets. For illustration, Table 7 demonstrates the variable selection from Logit model 4. In each of the 5 runs, new training and test sets were randomly selected and then applied to all four models. The training set consists of 85% of the tracks, while the resulting logit model is applied to the remaining 15% (test set). If the logit probability is >= 0.5, a song is classified as a hit. Hits are songs that reached a peak position in the top 10 or top 20. Non-Hits are songs whose peak position is in the 50-100 or 80-100 range. The models become more accurate the more chart positions are recorded and the higher the gap between hits and nonhits (Herremans et al., 2019). As expected, other studies whose non-hit group represented a selection of random songs, were able to predict Hit-Songs with a higher accuracy of 75% (Georgieva et al., 2018) and 88% (Middlebrook & Sheik, 2019).

Since songs with peak positions >100 (that is, positions 101, 102, etc.) presumably also take on characteristics of the Non-Hits in the range of 50-100 or 80-100, the Non-Hits in this study also represent all Non Chart entries.

Hit Accuracy is the proportion of correctly classified Hits to all actual Hits, while Non-Hit Accuracy is the proportion of correctly classified Non-Hits to all Non-Hits. Overall accuracy takes into account the absolute proportions of Hits and Non-Hits. Overall accuracy = (Hit-Accuracy*Hits + Non-Hit-Accuracy*Non-Hits)/(Hits + Non-Hits). Delta describes the difference between the Hit-Accuracy and the Non-Hit error rate (that are the wrongly as Hits classified Non-Hits). This means that a high delta is to be aimed for. It is noticeable that classification has an influence on Hit- and Non-Hit accuracy. Both models with a Top-20 Hit classification (Model 3 & 4) have a higher Hit-accuracy compared to the models with a Top-10 Hit classification (Model 1 & 2). These models with higher classification (Top-20) seem to classify generally more tracks as hits. And because the sets of these models contain more hits than Non-Hits, a higher total accuracy can be achieved (this also applies in the reverse case for the Non-Hits). Therefore, not only the overall accuracy, but also the delta should be considered. For me, despite lower overall accuracy, model 2 also seems suitable. Model 4 has the highest average accuracy of 69%.

Table 6 shows the logit model 4 after the feature selection of the first run. All variables have a P-value of <=0.3. The significance levels are declared as "*" for 5%, "**" for 1% and "***" for 0.1%. It is noticeable that more audio analysis variables (timbres) than audio features were retained. The reason for this disappearance may be that the audio features are calculated by other properties such as the timbres (Herremans et al., 2019). Overall, musical variables are strongly represented. This suggests that chart success can be increased just by musical song optimization. However, some socioeconomic variables are also relevant: As expected, we see those songs from “superstar artists” are more likely to be a Hit (with a highly significant z-value of 0.512), at least holding all other variables constant.

Table 6: Logit-Model Example

	(1)	
	hit	
hit		
timbre9_min	0.00703	(1.09)
timbre9_max	-0.00668	(-1.25)
1.key	0.368	(1.56)
2.key	0.474	(1.44)
8.key	0.327	(1.08)
timbre4_min	0.00315**	(2.59)
instrument~s	-1.479*	(-2.04)
timbre2_max	0.00204*	(2.15)
liveness	-1.946**	(-2.98)
superstar	0.512**	(3.01)
deep	-1.607***	(-5.43)
timbre10_s~v	0.128**	(2.82)
mode	-0.401*	(-2.44)
timbre8_min	-0.00444	(-1.42)
timbre12_max	-0.00846	(-1.15)
timbre3_mean	-0.00572	(-1.18)
timbre4_mean	0.0354***	(3.89)
trap	-0.746*	(-2.20)
timbre12_min	0.00922	(1.11)
timbre12_s~v	-0.109*	(-2.02)
timbre8_mean	-0.0240	(-1.56)
rnb	-1.278**	(-2.72)
timbre7_st~v	0.0386	(1.21)
section1_l~s	0.0282*	(2.08)
timbre2_min	0.00124	(1.29)
timbre2_st~v	0.0277**	(3.08)
timbre3_st~v	-0.0246*	(-2.16)
cloud	0.530	(1.16)
feature	0.317	(1.63)
timbre8_st~v	-0.133***	(-3.73)
_cons	3.149***	(3.44)
N	898	

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

References

Georgieva, E., Suta, M. & Burton, N. (2018). *HITPREDICT: PREDICTING HIT SONGS USING SPOTIFY DATA STANFORD COMPUTER SCIENCE 229: MACHINE LEARNING*. https://cs229.stanford.edu/proj2018/report/16.pdf

GfK Entertainment GmbH. (o. D.). *Wie werden die Charts erstellt?* Offizielle deutsche Charts. Abgerufen am 25. März 2022, von https://www.offiziellecharts.de/info/faq

Herremans, D., Martens, D. & Sörensen, K. (2014). Dance Hit Song Prediction. *Journal of New Music Research*, *43*(3), 291–302. https://doi.org/10.1080/09298215.2014.881888

Interiano, M., Kazemi, K., Wang, L., Yang, J., Yu, Z. & Komarova, N. L. (2018). Musical trends and predictability of success in contemporary songs in and out of the top charts. *R. Soc. open sci.*, *5*(5). https://doi.org/10.1098/rsos.171274

Leibenstein, H. (1950) Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand, reprinted in W. Breit and H.M. Hochman (eds.), Readings in Microeconomics, pp. 111–127. Holt, Rineheart and Winston, New York.

Mcdonald, G. (2021, September 24). *A short essay about long playlists of short tracks of rain noises and streaming-music economics.* Furia. Abgerufen am 25. März 2022, von https://www.furia.com/page.cgi?type=log#id466

Middlebrook, K. & Sheik, K. (2019, September). *SONG HIT PREDICTION: PREDICTING BILLBOARD HITS USING SPOTIFY DATA*. https://arxiv.org/pdf/1908.08609.pdf

Singhi, A. & Brown, D. G. (2014). Hit song detection using lyric features alone. *Proceedings of International Society for Music Information Retrieval*, *30*. https://ismir2014.ismir.net/LBD/LBD12.pdf

Strobl, E. A. & Tucker, C. (2000). The Dynamics of Chart Success in the U.K. Pre-Recorded Popular Music Industry. *Journal of Cultural Economics*, *24*, 113–134. https://doi.org/10.1023/A:1007601402245