Estimating Song Popularity on Spotify

Rohit Menon, Will Meyer, Alexander Lam, Scott Belarmino

Introduction

With global music industry revenue totaling $31.2 billion in 2022, there is much at stake for record executives, music producers, and musicians. While some artists struggle to make ends meet, the most popular artists have earned hundreds of millions and even billions of dollars. By increasing a song's popularity (output metric), all actors who created the song will gain from higher royalties, merchandise sales, and increased tour revenue.

Acknowledging the importance of a data-driven approach alongside creativity, this study employs observations from the Spotify API to estimate a song's popularity score, particularly focusing on the "pop" genre. Various song attributes (collectively denoted as our "X" concept) are analyzed to explore if a causal relationship exists between these variables and a song's popularity (our "Y" concept). The ultimate goal is to understand if manipulating these variables can increase the likelihood of a song becoming a 'hit', all else being equal.

For decades, artists and producers leveraged technological advancements to increase the loudness of their tracks, as producers believed that increasing volume could cause popularity. However, Spotify's normalization of track volumes potentially disrupted this trend, challenging producers to find alternative strategies for boosting song popularity on the platform. This study seeks to evaluate whether loudness may affect popularity in the modern era and if there are alternative methods to influence track popularity. We use multiple model specifications to measure these effects after additional controlling for potential confounding factors like song age and playlist subgenre.

Data and Methodology

To conduct our analysis on the impact of song characteristics on popularity, we utilized a dataset from Kaggle, comprising approximately 30,000 unique tracks sourced from the Spotify API. Since we are focused on estimating the relationship between song attributes and popularity, we sought to control for the significant variation of these characteristics across genres. We chose to narrow our scope to "pop" songs.

Our decision to limit our scope to pop was multifaceted. Primarily, we attempted to avoid the confounding variable of age - specifically, the "durability" of a song. For all other genres besides pop, we found a strong correlation between a track's popularity and the age of a song. Rap and EDM heavily favored newer songs, while the more popular rock songs tended to be older. In comparison, per Figure 1 below, track popularity varied much less compared to other genres and there is seemingly a small correlation between age and track popularity. Second, our exploratory data analysis found that pop had a broader variety of musical attributes than other genres, so we thought there might be more opportunities to exploit this variation to estimate the impact of these attributes on popularity.

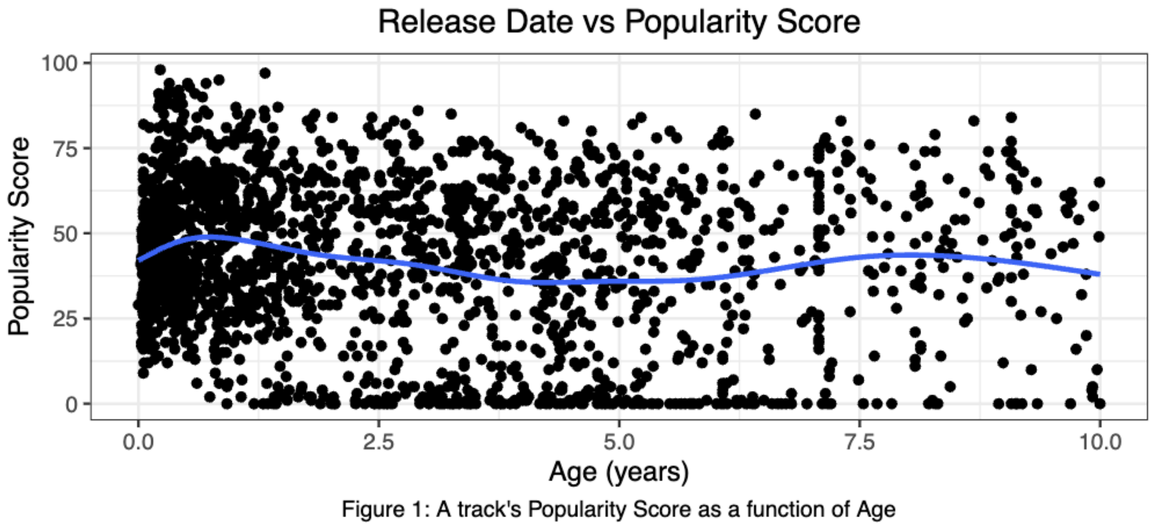

Figure 1: Found Pop exhibits an increase in popularity in the first ten years but has seemingly no correlation with release age.

This resulted in a dataset comprising 5,132 pop songs and involving 2,488 unique artists. To address potential issues of independent and identically distributed (IID) data, we randomly selected one song per artist, leaving 2,488 pop tracks in our data. This mitigates IID concerns of clustering by artist.

Our initial exploratory analysis seemed to confirm the findings of past music producers. That is, track loudness was positively correlated with track popularity (see Figure 2 below). However, one thing to note is that through Spotify's normalization process of loudness on their platform, the ideal loudness of a song on the app is -14 dB, and every user is able to select different levels from -23 to 0.

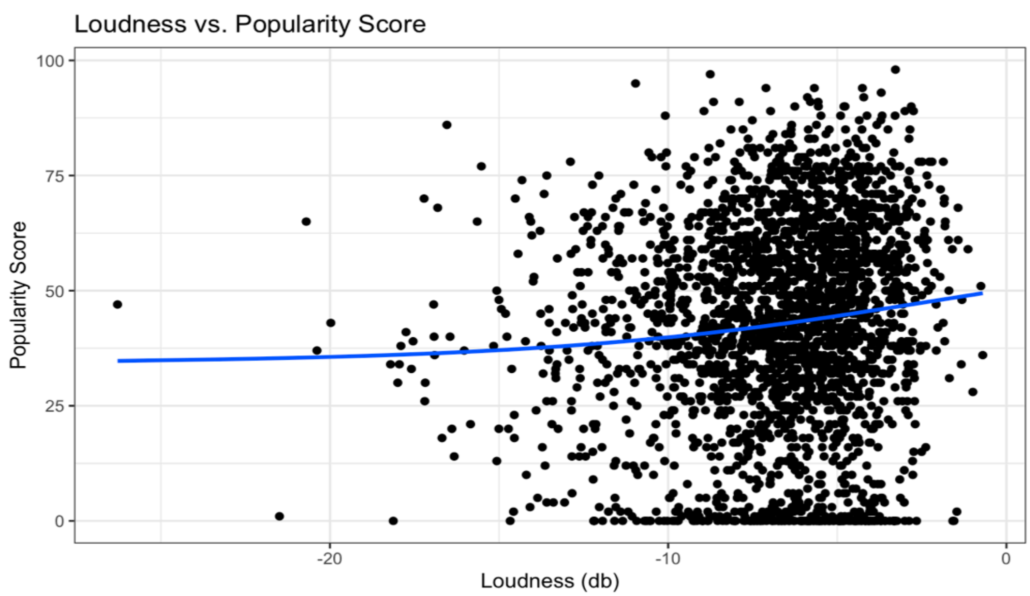

Figure 2: Loudness vs. Popularity Score

As we iteratively constructed our core model, we introduced fundamental music features like tempo, duration, and key, acknowledging the ongoing impact of loudness on a song's popularity. We employed a regression analysis to comprehend the interplay between loudness and different variables and their collective impact on the popularity of a song.

Each model we evaluated was of the following form:

- LATEX FORM: $\text{Popularity} = \beta_0 + \beta_1 \cdot \text{Loudness}$

where B1 represents the % increase in popularity for every 1% increase in loudness keeping other predictors constant.

Results

In Table 1, the results of 4 regression models that we conducted using popularity as our outcome variable are shown. We estimated the base model with only the loudness parameter in Model 1. In Model 2, we added an age parameter, which was calculated as the time between the date of the study and the track's release date. For Model 3, the parameters duration, tempo, mode, and key were built on top of the base model. The same parameters were applied for Model 4 with the addition of the "Sub Playlist Genre" variable.

In Model 1, our analysis revealed a statistically significant relationship, where a 1 unit increase in loudness corresponded to a notable 0.8 unit increase in the Spotify Popularity Score. However, it is crucial to note that loudness alone accounted for merely 1% of the total variability in the popularity score, indicating that loudness alone is just one aspect among many that contribute to a song's appeal on the platform.

The practical implication of this result is shown by the impact of a large change in loudness on the estimated popularity score. Specifically, a 10 dB increase or decrease in the loudness of a given song is associated with an approximate adjustment of 8 points in its popularity score. Since the songs in our sample generally range from -15 to -5 decibels, our model predicts at the top end of this range would have a popularity score nearly 10 points higher than a song towards the bottom of the decibel range.

In Model 2, aligning with the historical context of loudness wars, we might anticipate a positive covariance between the age of a song and loudness before Spotify's normalization efforts. Also, maybe some of the loudness effect is really a preference for newer songs, which also happen to be louder. However, our findings indicate that the influence of loudness remains robust even when accounting for song age in the model, with a beta value of 0.82. This suggests that the impact of loudness on the popularity score persists independently of song age, showcasing its significance in our model.

In Models 3 and 4, we broaden our investigation to explore how additional musical dimensions may contribute to predicting song popularity. Notably, the loudness parameter maintained its significance, with beta values of 0.78 and 0.45 across the two models. Similarly, the tempo parameter was significant, with beta values of -0.47 and -0.48. A noteworthy distinction between these models lies in the inclusion of the playlist sub-genre parameter. While this parameter predicts popularity less effectively, it yields a higher R-squared value of 0.08, suggesting that it captures a greater proportion of the variance in song popularity. In practical terms, this means excluding the playlist sub-genre may be more accurate in predicting the popularity of individual songs.

Table 1: Estimated Regressions

	Output Variable: Spotify Popularity Score			
	(1)	(2)	(3)	(4)
Loudness	0.80***	0.82***	0.78***	0.45**
	(0.16)	(0.16)	(0.17)	(0.16)
Age		0.0001	0.0004*	0.0003
		(0.0002)	(0.0002)	(0.0002)
Duration			−0.0000*	−0.0000
			(0.0000)	(0.0000)
Duration^2			0.00	0.00
			(0.00)	(0.00)
Tempo			−0.47***	−0.48***
			(0.14)	(0.14)
Tempo^2			0.002**	0.002**
			(0.001)	(0.001)
Mode			2.19*	2.46**
			(0.95)	(0.93)
Constant	47.84***	47.78***	92.40***	87.86***
	(1.18)	(1.18)	(11.23)	(11.31)
Key			✓	✓
Playlist Sub-Genre				✓
Observations	2,488	2,488	2,488	2,488
R^2	0.01	0.01	0.04	0.08
Residual Std. Error	22.73 (df = 2486)	22.73 (df = 2485)	22.48 (df = 2470)	21.99 (df = 2467)

Note: HC_1 robust standard errors in parentheses.

Limitations

8a. Statistical limitations of your Model

- *Make sure to evaluate all of the large sample Model assumptions (or the CLM if you have a small sample). However, you do not necessarily want to discuss every assumption in your report. Instead, highlight any assumption that might pose significant problems for your analysis. For any violations that you identify, describe the statistical consequences. If you are able to identify any strategies to mitigate the consequences, explain these strategies.*

8b. Structural limitations of your Model (i.e., reverse causality, omitted variables)

- Is there a possibility of reverse causality? If so, reason about the direction of bias this causes.
- Are there any outcome variables on the right-hand side? If so, reason about the direction of bias this causes.
- What are the most important omitted variables that you were not able to include? For each variable you name, reason about the direction of bias caused by omitting this variable and whether the omission of this variable calls into question the core results you are reporting.

A potential statistical limitation to our model was the number of pop artists with multiple songs in the data. This would potentially lead to clustering of the data based on the artist. To mitigate this, we decided only to select one song per artist, allowing for stronger reasoning for the independent and identically distributed (IID) assumption.

There could exist reverse causality between popularity and several of our selected variables. By choosing to focus on the “Pop” Genre, trends in popular songs could have a direct impact on the tempo, loudness, and mode of the produced song. As the name suggests, songs that are successful within “Pop” tend to follow the trends of the period in which they were released.

We also examined if there were any outcome variables found on the right-hand side of our equation. The variables were determined to be independent of each other. Thus, no outcome variables were detected.

Several omitted variables may influence our popularity score estimates. An example would be the marketing budgets attached to each track. If a track is backed with financial resources to purchase ad space on streaming platforms or incentivizes social media influencers to promote the song, we would assume this would have a positive correlation to the song’s popularity score. Marketing budgets would drive the estimates away from zero, thus overestimating the popularity score. Other omitted variables that may influence the effect include the artist’s popularity and social media presence.

Conclusion

This study provided a linear regression model to estimate the popularity score of a hit song on Spotify. From our base model, we found that with every decibel increase, the popularity score is estimated to increase by 0.80 points. Adding more variables to our model resulted in higher popularity scores driven by 0.78 points per decibel and a higher constant of 92.4 points.

We would like to deploy our model to aid record executives, music producers, and musicians in their goal to make a popular song. With additional variables and further regression analysis, our model may increase the accuracy of our popularity score estimate and provide value to the music industry.