Popularity Decoded: Unpacking the Association Between Track Audio Features and Popularity

University of California, Berkeley

Ryan Farhat-Sabet, Maia Kennedy, William Seward

Abstract

This study investigates the relationship between Spotify's "danceability" metric and track popularity using a dataset of over 30,000 songs. Leveraging proprietary Spotify audio features such as energy and instrumentalness, the analysis employs a baseline regression model to evaluate how danceability associates with popularity, complemented by multivariate models for additional insights. Initial findings suggest a statistically significant, though modest, relationship, with danceability explaining only a small fraction of the variance in track popularity. Challenges include data preparation, such as handling duplicates and addressing songs with zero popularity, as well as ensuring robust model assumptions. Despite limitations, the results highlight the complexity of musical popularity and the value of focusing on specific features like danceability for understanding listener engagement. This research offers implications for artists, record labels, and marketers seeking to optimize music for broader appeal.

Introduction

Spotify, a global leader in music streaming, has developed proprietary metrics such as danceability, instrumentalness, and energy to quantify and describe the features of songs. These sophisticated measures offer valuable insights into musical characteristics that influence listener engagement and preferences. By leveraging these tools, this study aims to address the question:

What is the association between danceability and track popularity?

This question may be compelling to a range of audiences, including artists looking to optimize their music for listener engagement, record labels seeking features that maximize success, or marketing professionals selecting music for campaigns. While many individuals have analyzed Spotify's audio features in the context of personal playlists or genre trends, scholarly research often focuses on broader models predicting music success using an array of audio features across genres. For example, prior studies have used machine learning to predict song popularity, often considering many features simultaneously. Our study narrows the scope by focusing on the association of a specific feature – danceability – with track popularity. By highlighting danceability, we aim to uncover patterns that connect musical features to commercial performance.

Description of the Data Source

The dataset, "30000 Spotify Songs," sourced from Kaggle and extracted via Spotify's API, contains information on over 30,000 tracks spanning various genres and release dates. This dataset was last updated on November 11th, 2023 and includes a comprehensive set of audio features, track identifiers, and song popularity with point-in-time values. Our dependent variable, **track_popularity**, is a proprietary Spotify-assigned score ranging from 0 to 100, reflecting listener engagement metrics such as recent play counts, listener engagement activity, current popularity (rather than historical success), and global versus local popularity. The primary independent variable in our baseline model is **danceability**, a measure from 0.0 to 1.0 of how suitable a track is for dancing based on tempo, rhythm, beat strength and overall regularity. Other features we consider include instrumentalness, which attempts to identify whether a song contains vocals, and energy, which measures how fast, loud, and noisy a particular song is based on perceived loudness, timbre, onset rate, and general entropy.

Data Wrangling

Scope the data: We ensured the dataset included only tracks with complete values for all relevant features. This dataset was very comprehensive and did not include any nulls or missing values for the variables of interest.

Handle duplicates: To prevent overcounting, we reviewed duplicate songs, defined as those with the same title and artist, regardless of release format (e.g., album or standalone track). Since our analysis focuses on audio features and popularity rather than publishing mode, we chose to address duplicates systematically. Interestingly, we discovered cases where duplicate songs had varying audio features and popularity metrics. To maintain consistency, we retained only the version with the highest popularity score, as we believe it best represents listener engagement and does not dilute the true popularity.

Consider zero popularity songs: After removing duplicated versions, we explored the remaining tracks that had a 0.0 popularity score. We examined several possibilities for this occurrence, including whether these songs truly had no listener engagement, if there was a reporting error with Spotify's API, or if Spotify's recommendation algorithm simply overlooked certain songs. We concluded that Spotify erroneously reporting 0.0 popularity would result in significant concern among artists, which justified retaining these songs with 0.0 popularity for further analysis. Ultimately, no additional data cleansing was required for this step.

Split the dataset: We split 30% of the data into an exploration set and the remaining 70% into a confirmation set at random.

Operationalization

Table 1: Variable Definitions

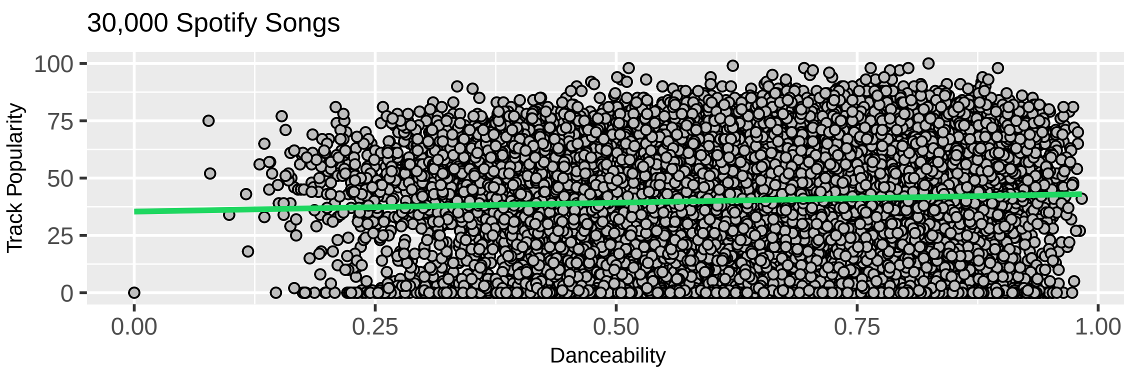
Variable	Feature	Definition	Scale
Dependent Variable(Y)	Track Popularity	A Spotify-assigned continuous metric based on listener engagement	0-100 (not popular → popular)
Independent Variable(X)	Danceability	A measure of how suitable a track is for dancing	0.0-1.0 (not danceable → danceable)
Additional X Variable	Instrumentalness	A score predicting the likelihood of a track being instrumental. Values greater than 0.5 are intended to represent instrumental tracks	0.0-1.0 (vocals → no vocals)
Additional X Variable	Energy	A perceptual measure of intensity and activity	0.0-1.0 (low energy → high energy)

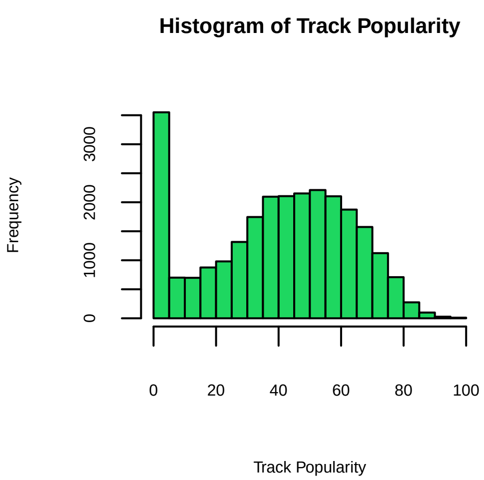
Danceability was chosen as the primary independent variable due to its perceived association with audience engagement and track popularity. The dataset initially contained 32,833 songs, which was reduced to 26,230 after removing duplicates.

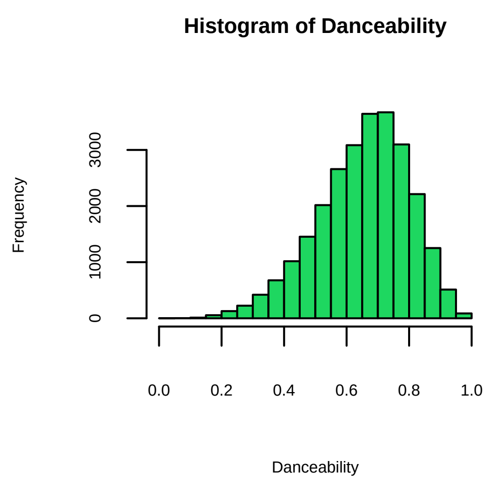
Visualization

For track popularity, we observe a relatively normal distribution, with many outliers of 0.0 popularity as previously mentioned. For danceability, we observe a left skewed distribution. For the joint distribution of danceability vs track popularity, we see a blob-like cloud; notice the 0.0 popularity songs hugging the X-axis.

```
`geom_smooth()` using formula = 'y ~ x'
```







Model Specification

Our study included a baseline and further comparison models to explore the associations between audio qualities and track popularity.

Baseline Model: lm(track_popularity~danceability)

We use an Ordinary Least Squares (OLS) regression model to estimate the relationship between danceability and track popularity. OLS regression is well-suited for this analysis due to its interpretability.

Comparison Model: lm(track_popularity~danceability+instrumentalness+energy)

We created multiple multivariate OLS regression models to explore additional audio features that may explain the variance of track popularity, including instrumentalness and energy, as well as other features from the Spotify API such as liveness, valence, key, and mode. We selected these variables to strengthen the robustness of the model and identify potential nonlinear relationships.

Model Assumptions

We investigated the two large model assumptions (IID and a unique BLP exists) to ensure our conclusions will be valid.

For IID, potential sources of dependence include scenarios where a hit song by an artist boosts the popularity of their other tracks. Additionally, unknown Spotify incentives should be considered, as the platform holds significant influence over music recommendations in ways that may not be fully transparent, potentially introducing biases that favor certain songs.

Potential non-identical distributions are also evident in our variables, as recent songs are more likely to be more popular than older tracks, and danceability systematically varies between genres (e.g., EDM vs. classical music). One way to possibly mitigate this is by stratifying the exploration and confirmation datasets. See appendix for more details.

To determine if a unique BLP exists, we start by checking whether a BLP exists at all. This involves examining the covariances between variables (e.g., Xi, Xj, and Xi, Y) to ensure there are no extreme values ("heavy tails"). Since all our variables are on a bounded scale, this condition is satisfied.

Next, we assess the uniqueness of the BLP by checking if any of the predictor variables (X) can be expressed as a linear combination of others. Our analysis confirms that there is no perfect or near-perfect collinearity, ensuring the BLP is unique.

Model Results and Interpretation

Regression Results

	Dependent variable:	
	Track Popularity w/ robust std errors	
	(1)	(2)
danceability	7.639***	6.536***
	(1.175)	(1.170)
instrumentalness		-13.611***
		(0.639)
energy		-12.310***
		(0.893)

```
Constant                    35.381***                  45.960***
                             (0.788)                    (1.035)

------------------------------------------------------------------------
Observations                 18,361                     18,361
R2                            0.002                      0.032
Adjusted R2                   0.002                      0.032
Residual Std. Error    23.168 (df = 18359)       22.826 (df = 18357)
F Statistic         42.597*** (df = 1; 18359) 200.500*** (df = 3; 18357)
========================================================================
Note:                                        *p<0.1; **p<0.05; ***p<0.01
```

Baseline Model

The OLS baseline model looks at track popularity as a function of danceability. When doing a simple linear regression, we can see that there is a statistically significant relationship between danceability and track popularity: for a 0.1 increase of danceability, there is a .7639 increase in track popularity.

However, our adjusted R^2 value is incredibly low; 0.2% of the variation in track popularity is explained by danceability in our model. This has nothing to do with the relationship we see between our variables, it just means that the amount of variation explained by our model is pretty low. This is exemplified by the large standard error among our residuals. This makes sense because track popularity is incredibly complex, and our model is a huge simplification. We are not accounting for so many other factors that influence popularity, so having a low R^2 here is to be expected.

Comparison Model

The comparison model looks at track popularity as a function of danceability, instrumentalness, and energy. We made sure to check for multicollinearity between all the variables, and there was very low correlation between all three of them. When regressing these variables against track popularity, we can see that there is a statistically significant relationship between track popularity and all three variables. For a 0.1 increase of danceability, there is a .6536 increase in track popularity. For a 0.1 increase of instrumentalness, there is a 1.3611 decrease in track popularity. For a 0.1 increase of energy, there is a 1.231 decrease in track popularity. Interestingly, as we have just highlighted, there is an inverse relationship between track popularity and instrumentalness.

Looking at our adjusted R^2 value, we see that 3.2% of the variation in track popularity is explained by our model. While this is over ten times better than our base model, it is still quite low, and we really see just how complex it is to model how popular a song is. Once again, our residuals have large standard errors, and we can see just how much data our model misses in the fitted values plot. When comparing against our base model, the plots look slightly less skewed because it is actually fitting better to the data ever so slightly. But we can still see how much variation exists here that can't be simplified and fitted by our model. The variables do explain a relationship with popularity here, but there is still so much that we are unable to capture with our data.

Opportunities for Future Research

If we pursued similar research in the future, we would be interested in further analyzing how the audio features themselves are associated with other audio features. We would also like to continue exploring how popularity of tracks changes over time. Lastly, we are interested in testing whether transforming continuous measures, such as danceability, into binary variables (danceable vs. not danceable with a 0.5 cutoff) could enhance the performance of our models.

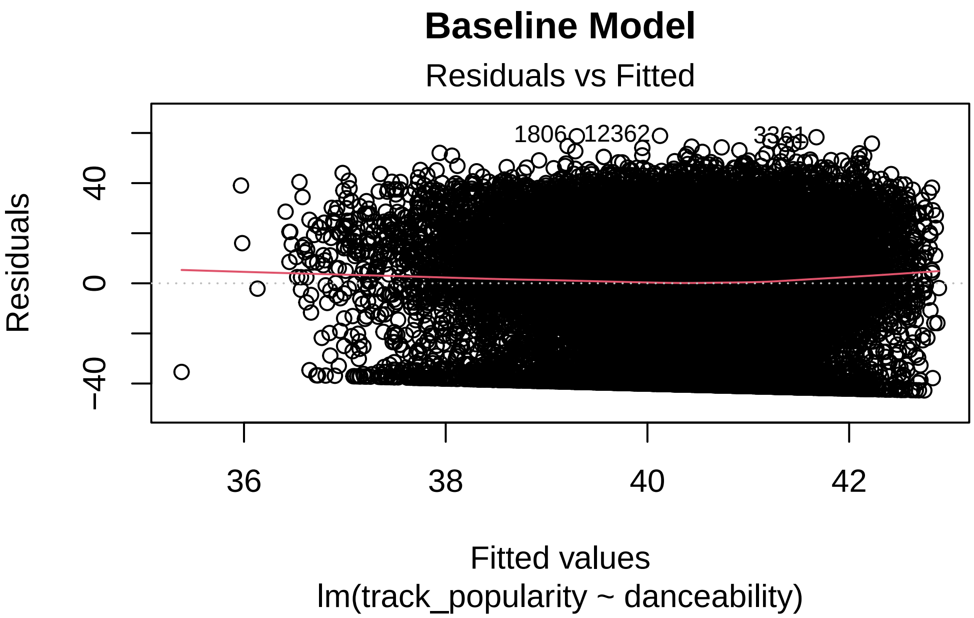
Appendix

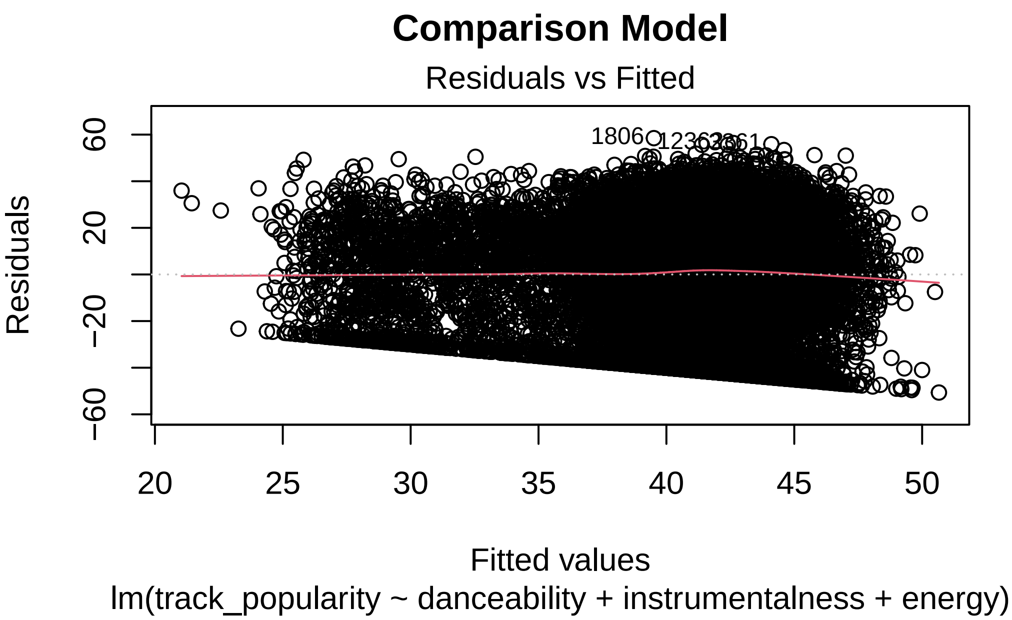
- **Data Source:** https://www.kaggle.com/datasets/joebeachcapital/30000-spotify-songs/data
- **Additional Models Explored:**
  - Regressing mode on danceability does not meaningly improve R^2
  - Regressing duration on danceability actually reduces R^2 and is not very interpretable despite being statistically significant
  - Regressing tempo on danceability does not meaningly improve R^2
  - Regressing valence on danceability shows that valence is not a statistically significant variable
  - Regressing loudness on danceability does not meaningly improve R^2

Regression Results

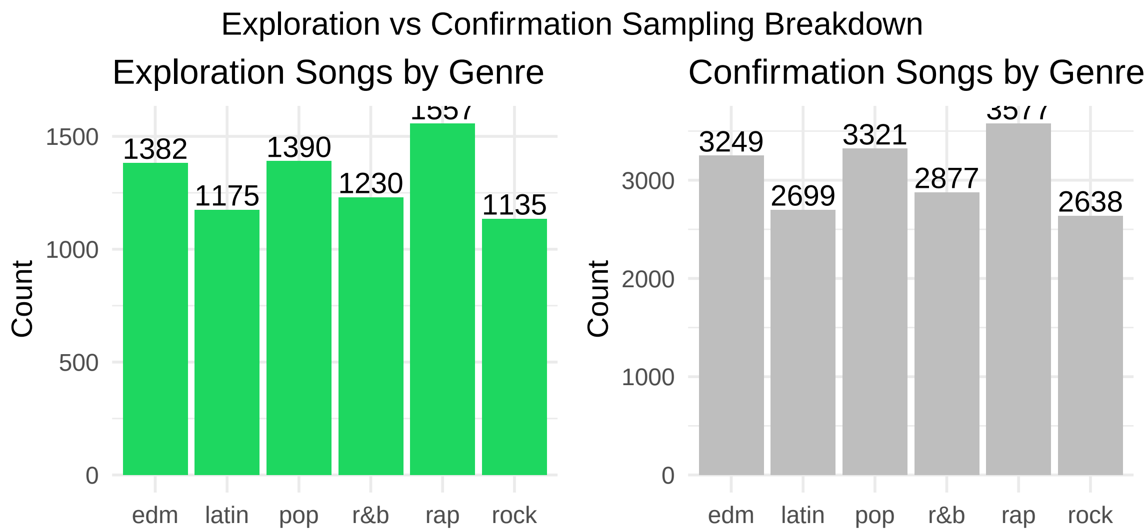
	Dependent variable:						
	Track Popularity w/ robust std errors						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
danceability	7.639***	8.103***	5.995***	8.707***	6.901***	7.880***	6.536***
	(1.175)	(1.825)	(1.823)	(1.847)	(1.946)	(1.826)	(1.170)
mode		1.253**					
		(0.529)					
duration_ms			-0.00005***				
			(0.00000)				
tempo				0.024**			
				(0.010)			
valence					1.977		
					(1.209)		
loudness						0.315***	
						(0.085)	
instrumentalness							-13.611***
							(0.639)
energy							-12.310***
							(0.893)
Constant	35.381***	34.614***	47.949***	32.057***	35.107***	37.600***	45.960***
	(0.788)	(1.268)	(1.619)	(1.866)	(1.241)	(1.352)	(1.035)
Observations	18,361	7,869	7,869	7,869	7,869	7,869	18,361
R2	0.002	0.003	0.019	0.003	0.003	0.004	0.032
Adjusted R2	0.002	0.003	0.019	0.003	0.003	0.004	0.032
Residual Std. Error	23.168 (df = 18359)	23.283 (df = 7866)	23.095 (df = 7866)	23.283 (df = 7866)	23.287 (df = 7866)	23.272 (df = 7866)	22.826 (df = 18357)
F Statistic	42.597*** (df = 1; 18359)	12.388*** (df = 2; 7866)	76.674*** (df = 2; 7866)	12.375*** (df = 2; 7866)	10.970*** (df = 2; 7866)	16.191*** (df = 2; 7866)	200.500*** (df = 3; 18357)
Note:							*p<0.1; **p<0.05; ***p<0.01

- **Additional features of the dataset:** There are 12 audio features for each track, including confidence measures like acousticness, liveness, speechiness and instrumentalness, perceptual measures like energy, loudness, danceability and valence (positiveness), and descriptors like duration, tempo, key, and mode (https://www.kaylinpavlik.com/classifying-songs-genres/)
- **Residuals vs. Fitted**





- **Exploration vs. Confirmation Sampling Breakdown**


Exploration vs Confirmation Sampling Breakdown
Exploration Songs by Genre
Confirmation Songs by Genre
Count
1500
1000
500
0
1382
1175
1390
1230
1135
Count
3000
2000
1000
0
3249
2699
3321
2877
2638
edm
latin
pop
r&b
rap
rock
edm
latin
pop
r&b
rap
rock