TEAM TOKYO

Popularity Scientists

2025

SPOTIFY TOP50

THE SCIENCE BEHIND ENTERING THE CHARTS

OUR TEAM

IDENTIFYING WHAT MAKES YOU FAMOUS

THE SINGER

THE SQL SAVIOR

THE MUSIC INDUSTRY EXPERT

THE DATA FARMER

INTRO

WHAT MAKES SONGS BREAK INTO THE CHARTS?

Music streaming is in our everyday life, but few of us understand how things work. Why is a song popular, what makes it stand out. Most of the factors belong to the subconscious, the unperceivable.

Today we are going to examine Spotify's top50 charts to identify which patterns make a song stand out. To help with that, Spotify provides a database with attributes that can help us reach a conclusion.

OUR MISSION

PROBLEM STATEMENT

While all songs in the Spotify Top 50 are successful, there is still significant variation in their popularity scores and chart positions. Music labels and producers want to understand: What musical characteristics separate the biggest hits (positions 1-20) from moderate hits (positions 20-50)?
By analyzing the Top 50 songs of 2019, we aim to identify which features (energy, danceability, valence, loudness) are most strongly associated with reaching the very top of the charts. This will help producers optimize their songs to maximize hit potential.

HYPOTHESIS

1. **CERTAIN GENRES DOMINATE THE TOP 10 WHILE OTHERS APPEAR ONLY IN LOWER POSITIONS**

2. **TOP 20 SONGS HAVE MORE POSITIVE MOOD (VALENCE) THAN LOWER-RANKED SONGS**

3. **TOP 20 SONGS ARE LOUDER THAN SONGS IN POSITIONS 21-50**

4. **TOP 20 SONGS HAVE SIGNIFICANTLY HIGHER ENERGY AND DANCEABILITY THAN POSITIONS 21-50**

5. **SONGS WITH LYRICS RANK HIGHER THAN PURELY INSTRUMENTAL**

6. **COLLAB SONGS RANK HIGHER THAN SINGLE ARTIST RELEASES**

WHAT IS SPOTIFY'S TOP50

A COMPILATION OF THE MOST STREAMED SONGS ON THE PLATFORM OF SPOTIFY, AGNOSTIC OF GENRE, OR ARTIST.

SONGS MAKES IT TO THE TOP50, BASED ON ONE METRIC THAT IS CALLED :

POPULARITY

THE ATTRIBUTES OF A SONG

GENRE

The style of music the songs belongs into

DANCEABILITY

How suitable the song is for dancing based on tempo, rhythm, and beat.

VALENCE

The musical positivity or happiness of the song.

SPEECHINESS

The amount of spoken words or rap in the song.

BPM

The speed of the song measured in beats per minute.

LOUDNESS

The overall volume of the song in decibels.

LENGHT

The duration of the song in milliseconds.

COLLABORATION

How well-known or widely listened to the song is.

ENERGY

How intense and dynamic the song feels.

LIVENESS

The likelihood that the song was recorded live.

ACCOUSTICNESS

How acoustic or non-electronic the song sounds.

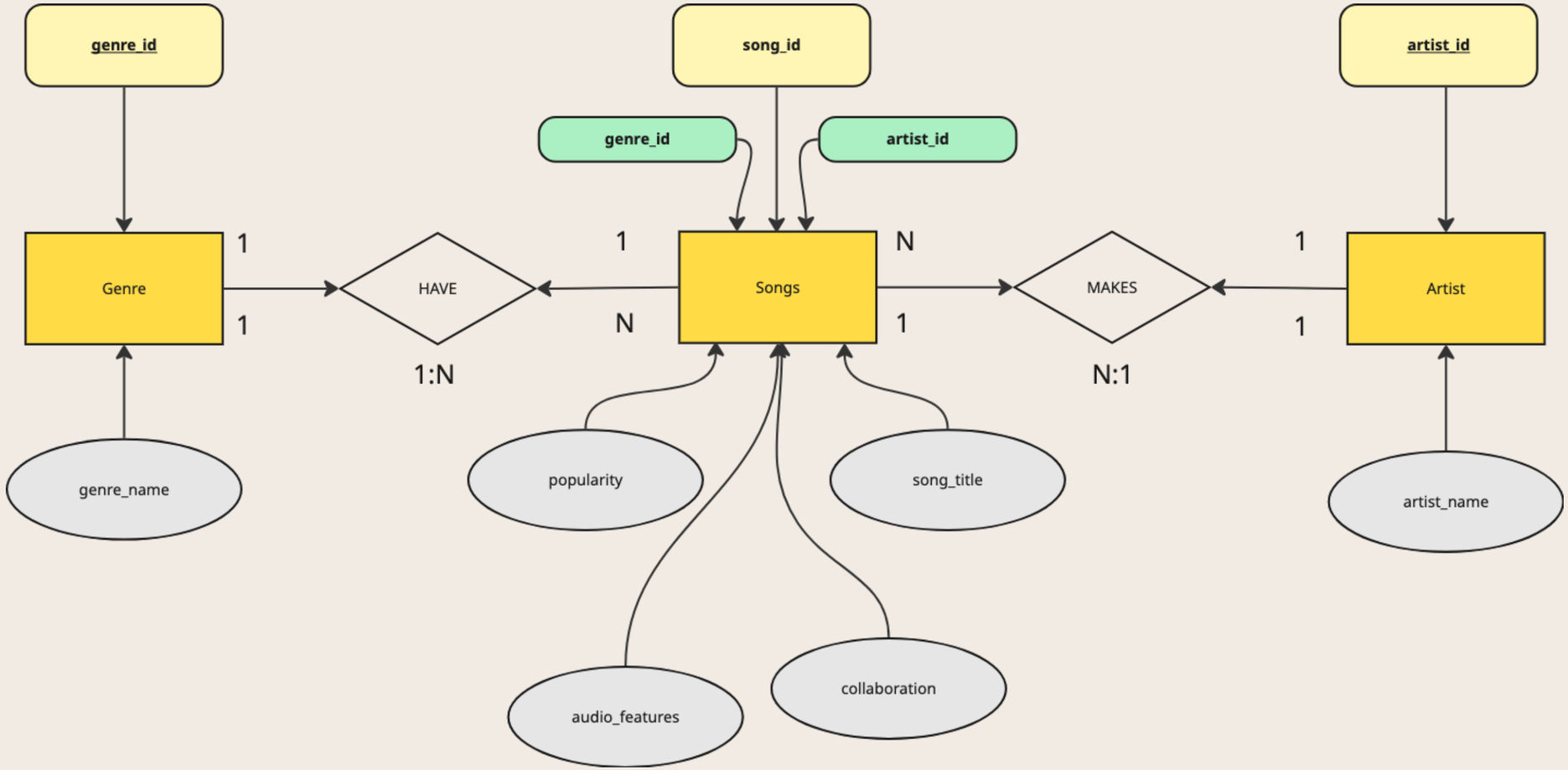
POPULARITY

Whether the song features multiple artists working together.

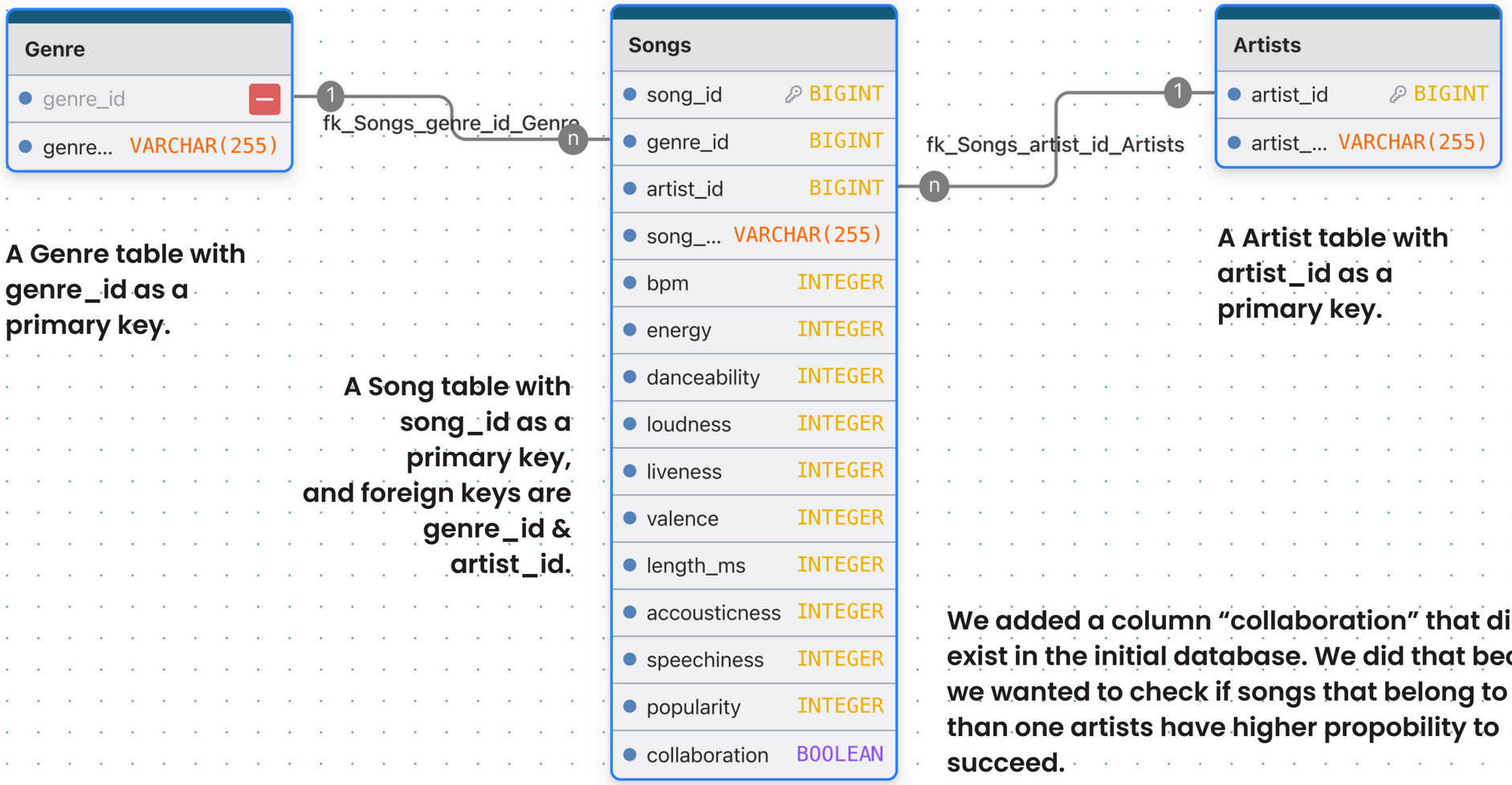
SO THE QUESTIONS IS

WHICH ATTRIBUTES DRIVE POPULARITY?

OUR CONCEPTUAL DIAGRAM



AND HOW OUR ERD LOOKS



A Genre table with genre_id as a primary key.

A Song table with song_id as a primary key, and foreign keys are genre_id & artist_id.

A Artist table with artist_id as a primary key.

We added a column “collaboration” that did not exist in the initial database. We did that because we wanted to check if songs that belong to more than one artists have higher propobility to succeed.

METHODS

OF ANALYSIS WE USED TO IDENTIFY PATTERNS AND CORRELATIONS

1.

REGRESSION (INFERENCE TYPE) **ANALYSIS**

A statistical method that models the relationship between a dependent variable and one or more independent variables. It estimates how changes in predictors are associated with changes in the outcome.

2.

COORRELATION MATRIX

A table showing pairwise correlation coefficients between variables. It summarizes the strength and direction of linear associations without implying causation or controlling for other variables.

3.

VARIANCE INFLATION FACTORS (OR VIF)

Variance Inflation Factor quantifies how strongly a predictor is linearly related to other predictors in a regression model. High VIF means high multicollinearity, making coefficient estimates unstable and unreliable.

REGRESSION ANALYSIS

OUTCOME

OLS Regression Results

Dep. Variable:	popularity	R-squared:	0.381
Model:	OLS	Adj. R-squared:	[illegible]
Method:	Least Squares	F-statistic:	1.704
Date:	Thu, 04 Dec 2025	Prob (F-statistic):	0.102
Time:	11:38:58	Log-Likelihood:	-133.56
No. Observations:	50	AIC:	295.1
Df Residuals:	36	BIC:	321.9
Df Model:	13		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	90.0372	9.416	9.562	0.000	70.940	109.134
bpm	0.0104	0.026	0.407	0.687	-0.041	0.062
energy	0.0097	0.070	0.139	0.890	-0.132	0.151
danceability	-0.0214	0.057	-0.377	0.709	-0.137	0.094
loudness	-0.2474	0.446	-0.555	0.582	-1.152	0.657
liveness	0.0595	0.059	1.001	0.324	-0.061	0.180
valence	-0.0574	0.033	-1.715	0.095	-0.125	0.010
length_ms	-0.0136	0.019	-0.703	0.487	-0.053	0.026
accousticness	-0.0189	0.037	-0.516	0.609	-0.093	0.055
speechiness	0.0668	0.076	0.880	0.385	-0.087	0.221
genre_hip hop	1.2672	2.971	0.427	0.672	-4.758	7.292
genre_latin	3.2321	2.713	1.191	0.241	-2.270	8.735
genre_pop	-1.5777	2.174	-0.726	0.473	-5.987	2.832
genre_rap	3.0738	2.923	1.051	0.300	-2.855	9.003

Omnibus:	10.280	Durbin-Watson:	2.136
Prob(Omnibus):	0.006	Jarque-Bera (JB):	11.466
Skew:	-0.770	Prob(JB):	0.00324
Kurtosis:	4.769	Cond. No.	4.26e+03

WHAT TO KEEP FROM HERE:

R-SQUARED: 0.381 << 1

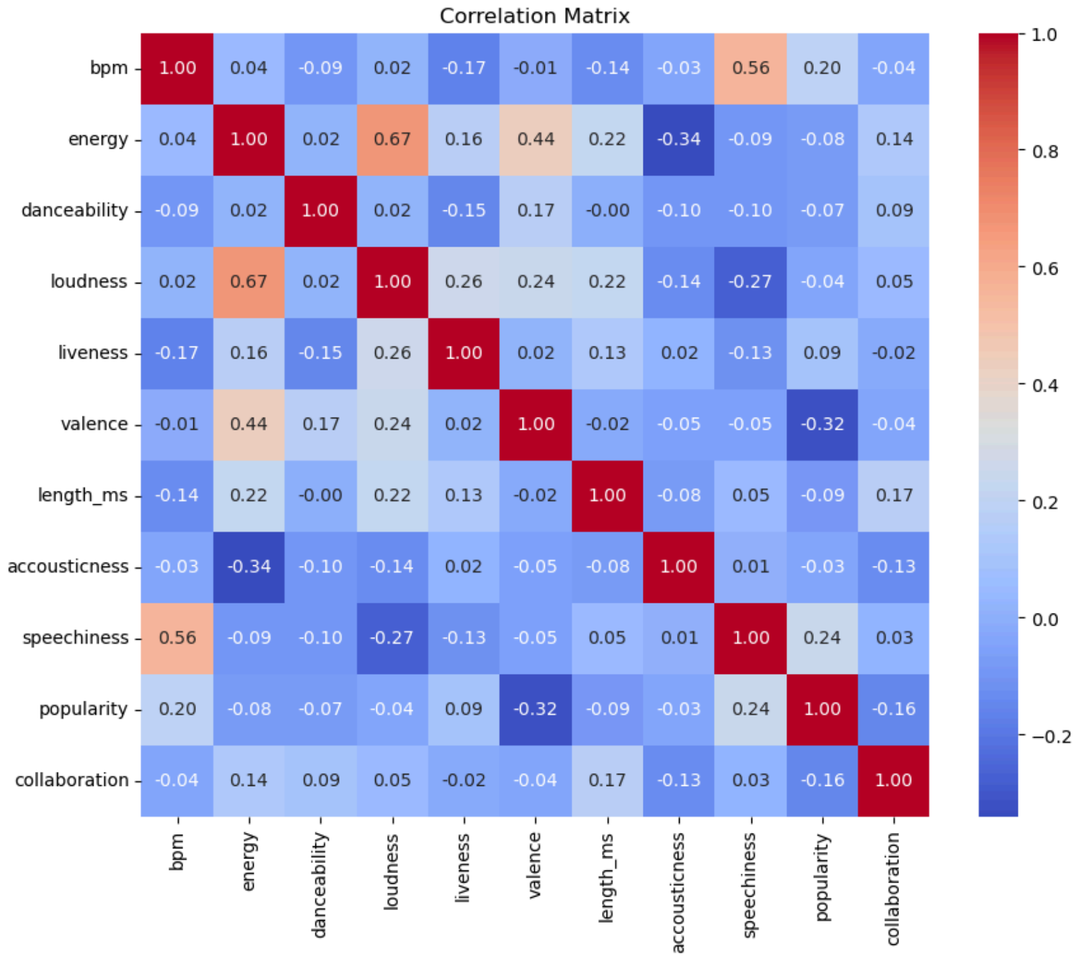
P>|T| VALUES < 0.8

Nothing in these results supports a strong claim about which attributes drive popularity.

The coefficients do not stabilize; significance is absent; the model complexity is too high relative to sample size.

CORRELATION ANALYSIS

OUTCOME



WHAT TO KEEP FROM HERE:

−0.3 < CORRELATION < +0.2.

In practical terms: none of these audio features linearly explain popularity in a meaningful way.

If we attempted to keep something, though more as a speculation, that could be that:

SAD SONGS HIT HARDER

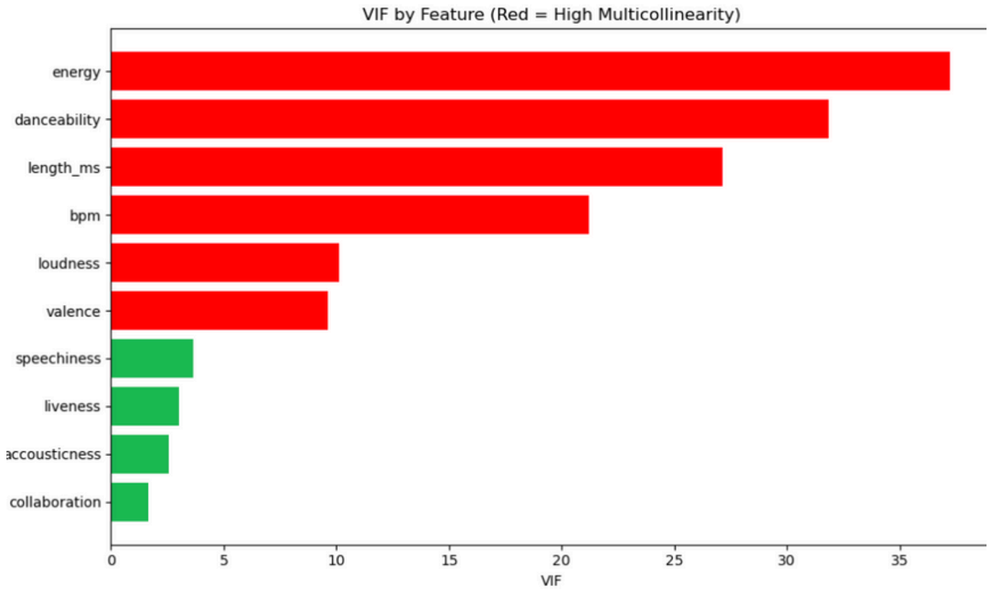
Happiness (valence) has a slightly negative correlation to popularity.

ORGANIC VS. BEATS

Acoustic songs, involving guitars and pianos, appear to be more popular than those with electronic beats.

VARIANCE INFLATION FACTORS

OUTCOME



WHAT TO KEEP FROM HERE:

VIF > 10, WE CAN FORGET ABOUT IT

Attributes energy (37), danceability (32), length_ms (27), bpm (21) These are extremely collinear.

These variables move together so strongly that in a regression they cannot be given separate, stable coefficients.

Attribute valence (9.6)

Borderline problematic. They also overlap with other features.

Attributes liveness, accousticness, speechiness, collaboration (VIF < 5)

Those are the only reasonably independent predictors

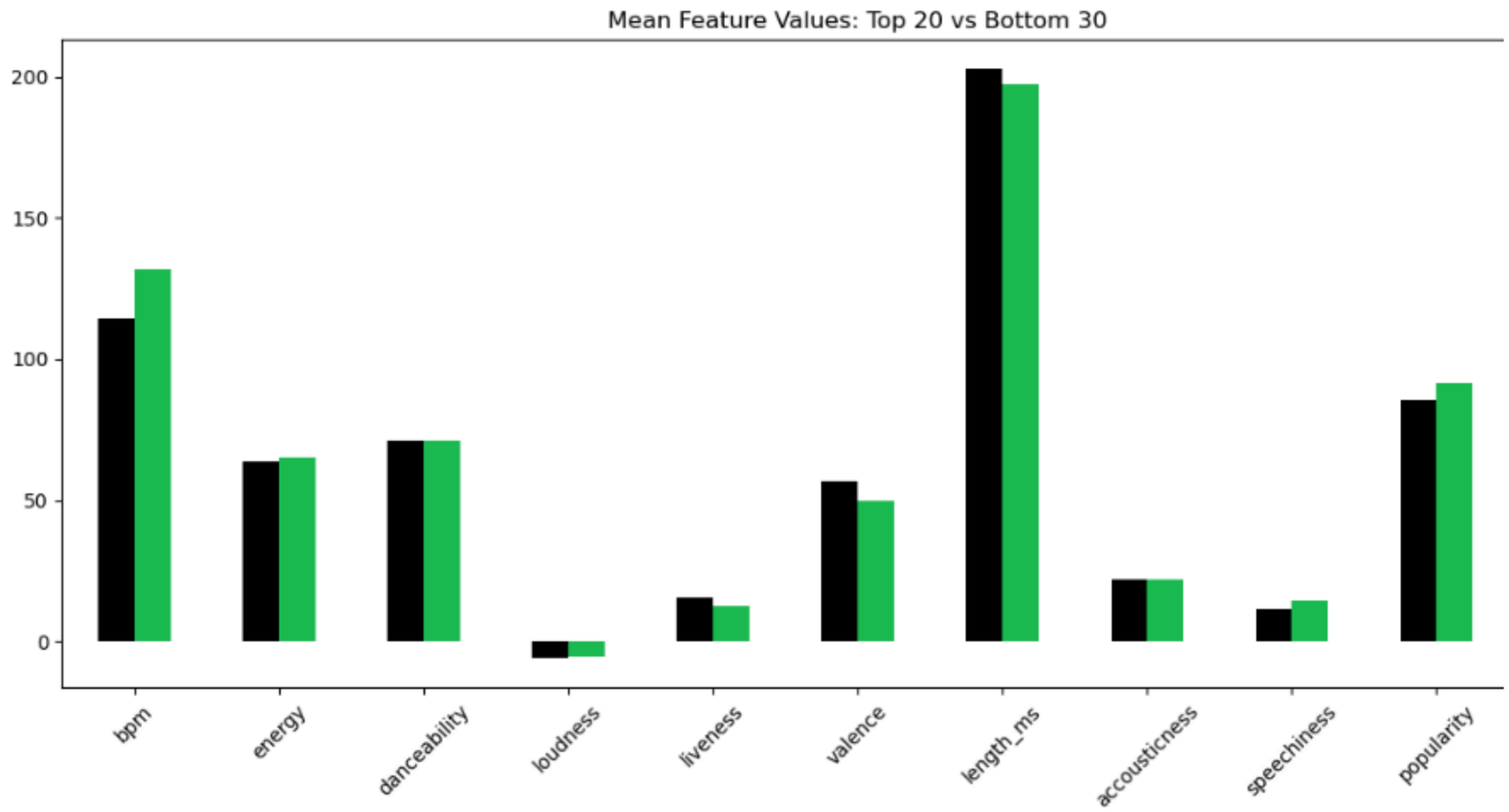
HYPOTHESIS REMINDER

OUR HYPOTHESIS 2, 3, 4, AND 5 THEORIZED THAT THE TOP 20 SONGS ARE:

- **HAPPIER IN MOOD (VALENCE)**
- **LOUDER**
- **MORE ENERGETIC**
- **MORE SPEECHY (HAVE MORE HUMAN SINGING)**

TOP20 VS. REST30

OUTCOME



THERE IS VIRTUALLY NO DIFFERENCE IN ANY ATTRIBUTE BETWEEN TOP20 AND REST30 SONGS

NOW THAT WE ARE COMPLETEE WITH A WIDE RANGE STATISTICAL ANALYSIS OF OUR DATA SET

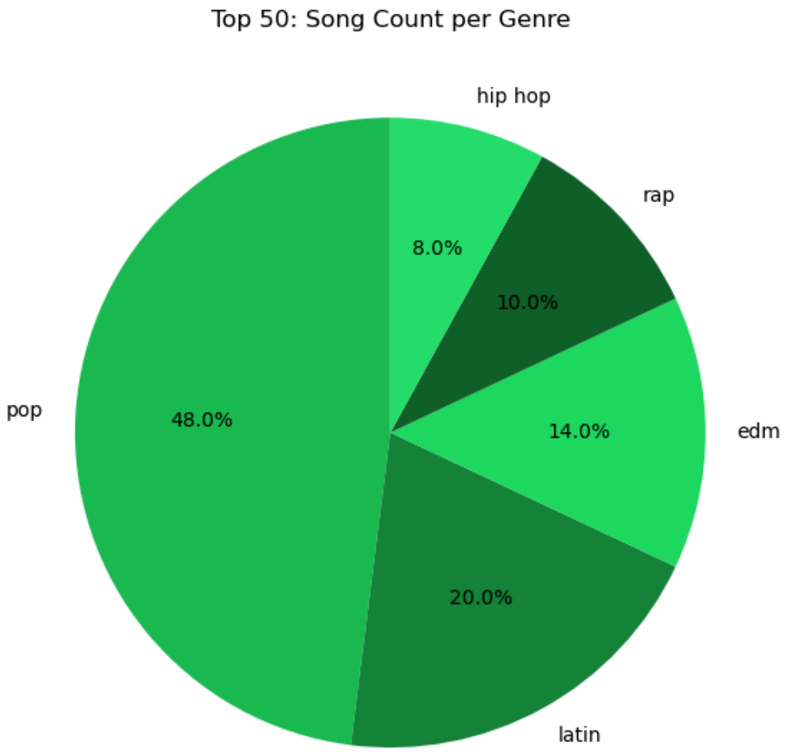
LET'S CHECK OUR REMAINING HYPOTHESIS

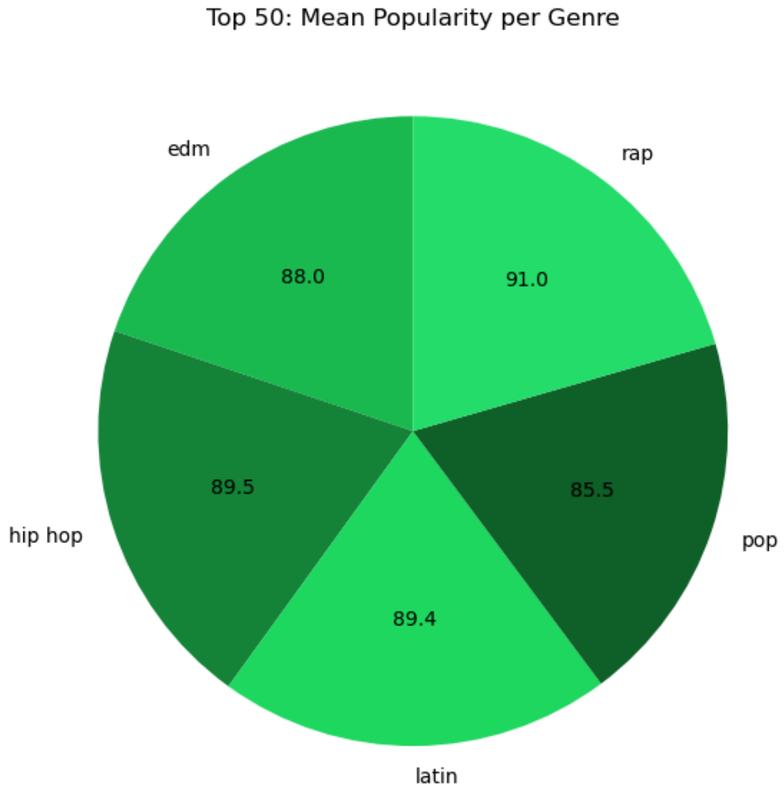
1. CERTAIN GENRES DOMINATE THE TOP 20 WHILE OTHERS APPEAR ONLY IN LOWER POSITIONS

6. COLLAB SONGS RANK HIGHER THAN SINGLE ARTIST RELEASES

GENRES & POPULARITY

OUTCOME





WHAT DO WE SEE:

1.

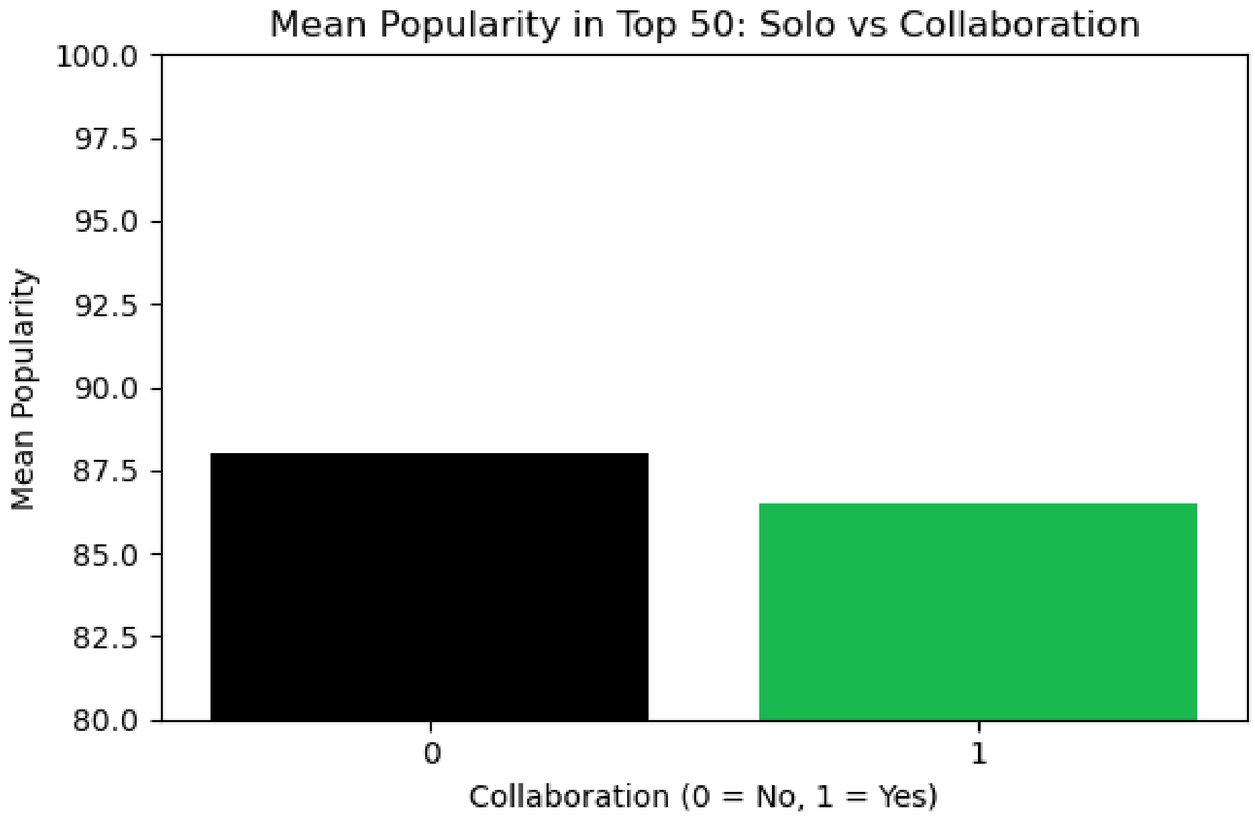
SOME GENRES APPEAR MORE FREQUENTLY IN THE TOP50, WHILE WE DON'T SEE GENRES LIKE ROCK, METAL, TECHNO, AND OTHERS MAKING IT IN THE TOP50.

2.

THE GENRE ITSELF HOWEVER, DOES NOT SEEM TO AFFECT POPULARITY OF A SONG, WITHIN THE DATASET. THAT COULD MEAN, THAT THE ONLY REASON THAT MORE POP SONGS ARE PRESENT THAN EDM, IS DUE TO THE SHEER AMOUNT OF POP SONGS RELEASED, COMPARED TO EDM.

COLLABS & POPULARITY

OUTCOME



WHAT DO WE SEE:

1.

AN ARTIST COLLABORATION SEEMS TO BE A NEGLIGIBLE FACTOR FOR SONG DIFFERENTIATION IN THE TOP50

2.

HOWEVER, IT MIGHT BE THE REASON IT MADE IT TO THE TOP50 AT THE FIRST PLACE. IN ORDER FOR US TO JUDGE BETTER, WE NEED TO COMPARE AGAINST LOWER IN POPULARITY SONGS.

1

HYPOTHESIS NOT CONFIRMED

None of our hypothesis were proven as statistically significant.

2

SMALL SAMPLE SIZE

That is normal. With a sample size of only 50 songs, variability in popularity is small.

3

EXTENTION NEEDED

To be able to dive deeper, we will need to merge a bigger database to the existing, bringing the amount of songs to 1000.

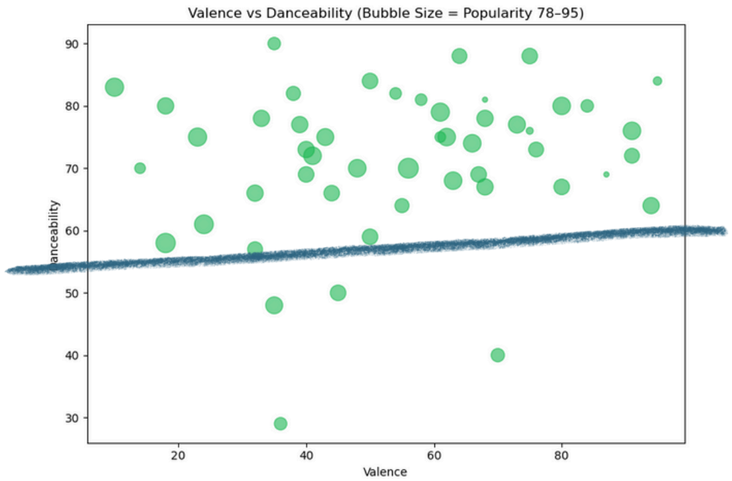
MUSIC IN THE AGE OF STREAMING

ALL IN ALL

EXTRA ANALYSIS

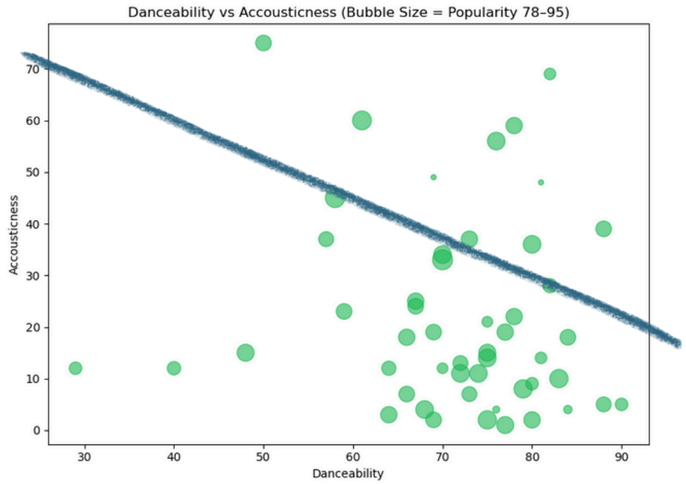
FOR FUN

WE DANCE MORE TO HAPPIER SONGS



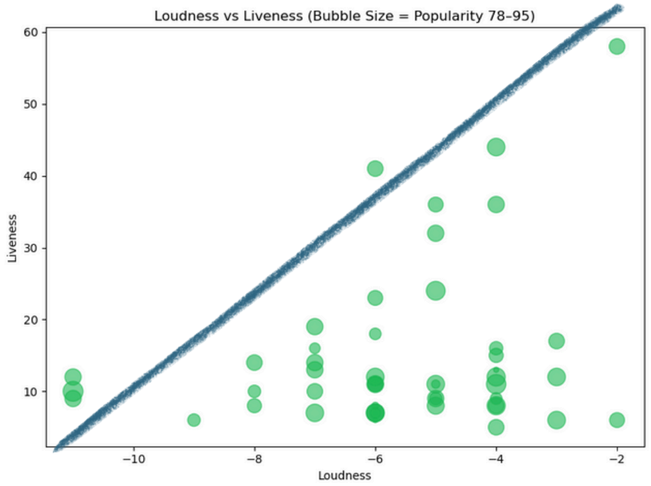
NOT

ELECTRONIC SONGS ARE MORE DANCABLE



KINDA

LOUD SONGS ARE PERCEIVED MORE LIVELY



YES

FROM POPULARITY SCIENTISTS

THANK YOU

thepopularityscientists.com