Tuning into Popularity: Predicting Spotify Track Success Through Audio Features

By Victoria Brigola | 2025

What's the Issue...

Over **100,000** tracks are uploaded to Spotify **EVERY DAY.**

Manual curation **CANNOT** scale to this volume, leading to **MISSED OPPORTUNITIES** for rising hits.

Solution...

Building a machine learning model that **predicts track popularity using audio features**, no streaming data required!

BUT CAN THIS BE DONE?

Let's find out if my model can predict a hit...

Meet the Features

1. I selected Spotify datasets with rich audio features and merged them using identifiers.

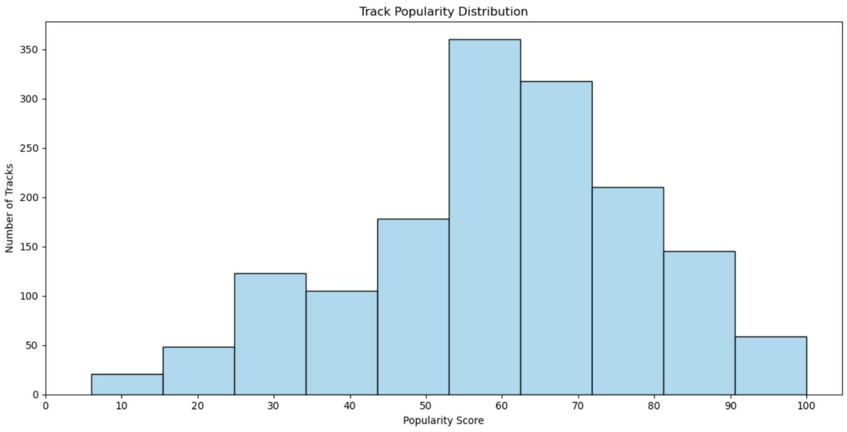
2. Focusing on key attributes like **danceability, loudness, valence and energy** that align with track popularity.

3. Each track is described by 22 features; popularity is scored from 0 to 100.

danceability	...	energy	instrumentalness	key	liveness	loudness	mode	speechiness	tempo	time_signature	valence
0.679	...	0.770	0.000000	0.0	0.0825	-3.537	1.0	0.1900	161.721	4.0	0.839

Below is a sample of the audio feature columns used for modeling and each row represents a single track.

What's POPULAR, really?



This histogram shows how track popularity scores are distributed across the dataset, with most songs falling in the mid-range and fewer reaching the top-tier popularity.

★ I removed tracks that scored 0 popularity to reduce the noise in the distribution.

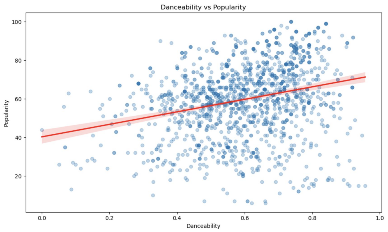
★ Most tracks cluster between 40-70, with fewer reaching high popularity.

★ The distribution is heavily right-skewed, with very few tracks approaching 90+.

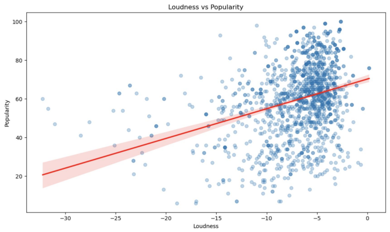
★ This imbalance helped justify binning popularity into categories (Low, Medium, and High) for classification.

Features that Hit Different

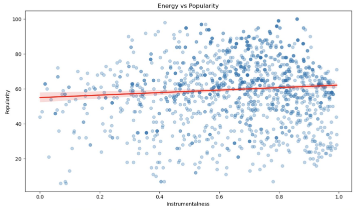
Danceability shows the strongest positive trend. The more danceable tracks are the more likely to chart!



Loudness has a wider spread but leans positively. Popular tracks tend to be louder overall.



Energy and valence show weaker but still visible correlations with higher popularity scores.



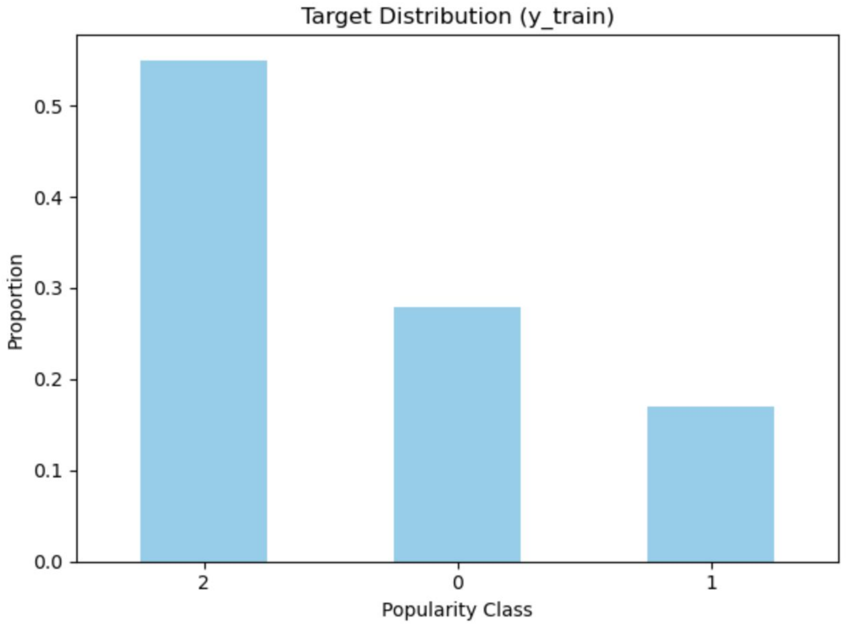
While no single variable is dominant alone, together they form a meaningful prediction base.

Finding the Right Fit

- ★ I initially explored the problem as a regression task, predicting continuous scores.
- ★ Even after trying regularization techniques, results were weak and inconsistent.
- ★ While early regression plots (ex: regplots in 'Features that Hit Different') revealed weak but positive trends, the relationships weren't strong or linear enough to support accurate prediction leading to my shift towards classification methods.

Classifying the Hits

- ★ Popularity scores were binned into Low (Class 0), Medium (Class 1), High (Class 2)
- ★ Used LabelEncoder to convert class labels to integers.
- ★ Applied Standard Scaler to normalize all numeric audio features.
- ★ Split the dataset into 80% training and 20% for model evaluation.



Battle of the Classifiers

Cross-Validation Results:

	F1 Macro
Tuned RF	0.715193
Bagging	0.715167
Random Forest	0.702979
DT Gini	0.669170
DT Entropy	0.668360
Logistic Regression	0.543641

- ★ Models tested: Decision Tree (Gini & Entropy), Random Forest, Bagging, and Logistic Regression.
- ★ Applied cross-validation to evaluate consistency and generalization.
- ★ Bagging and Tuned Random Forest were the top 2 on F1 macro scores (~0.71).

The Drop: Final Model Performance Results

Bagging Classifier was the most balanced model, achieving strong F1 macro Scores across most popularity classes.

The confusion matrix revealed Class 1 (medium popularity) had the highest misclassification due to class imbalance.

Overall, the model generalized well on the test set, making it a strong candidate for real-world deployment.

So, did my model work...

Answer is YES, my final model was able to predict track popularity with consistent accuracy using only audio features like danceability, energy, loudness, and valence.

Just the sound and the data behind it.