

NOIZE

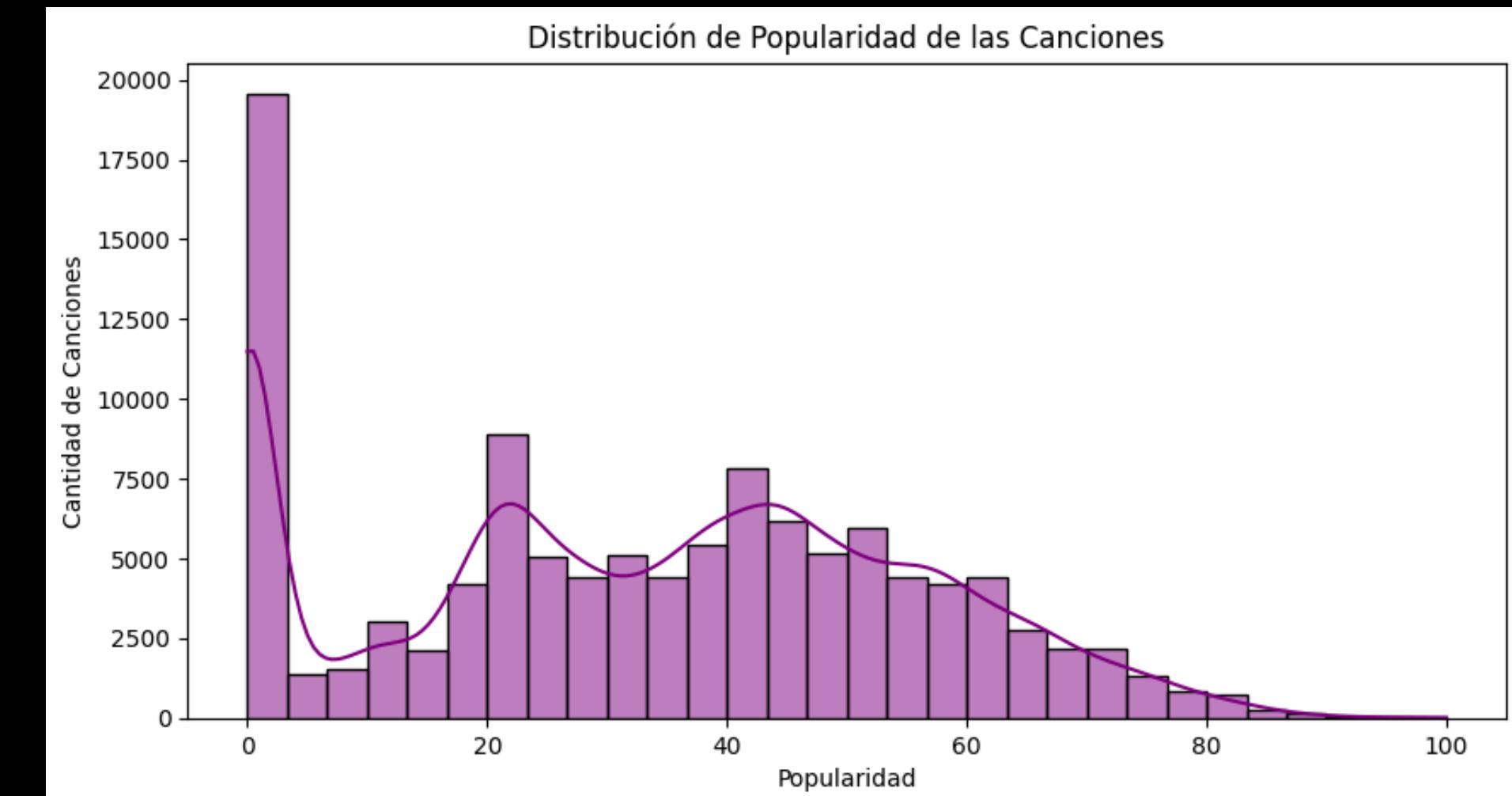


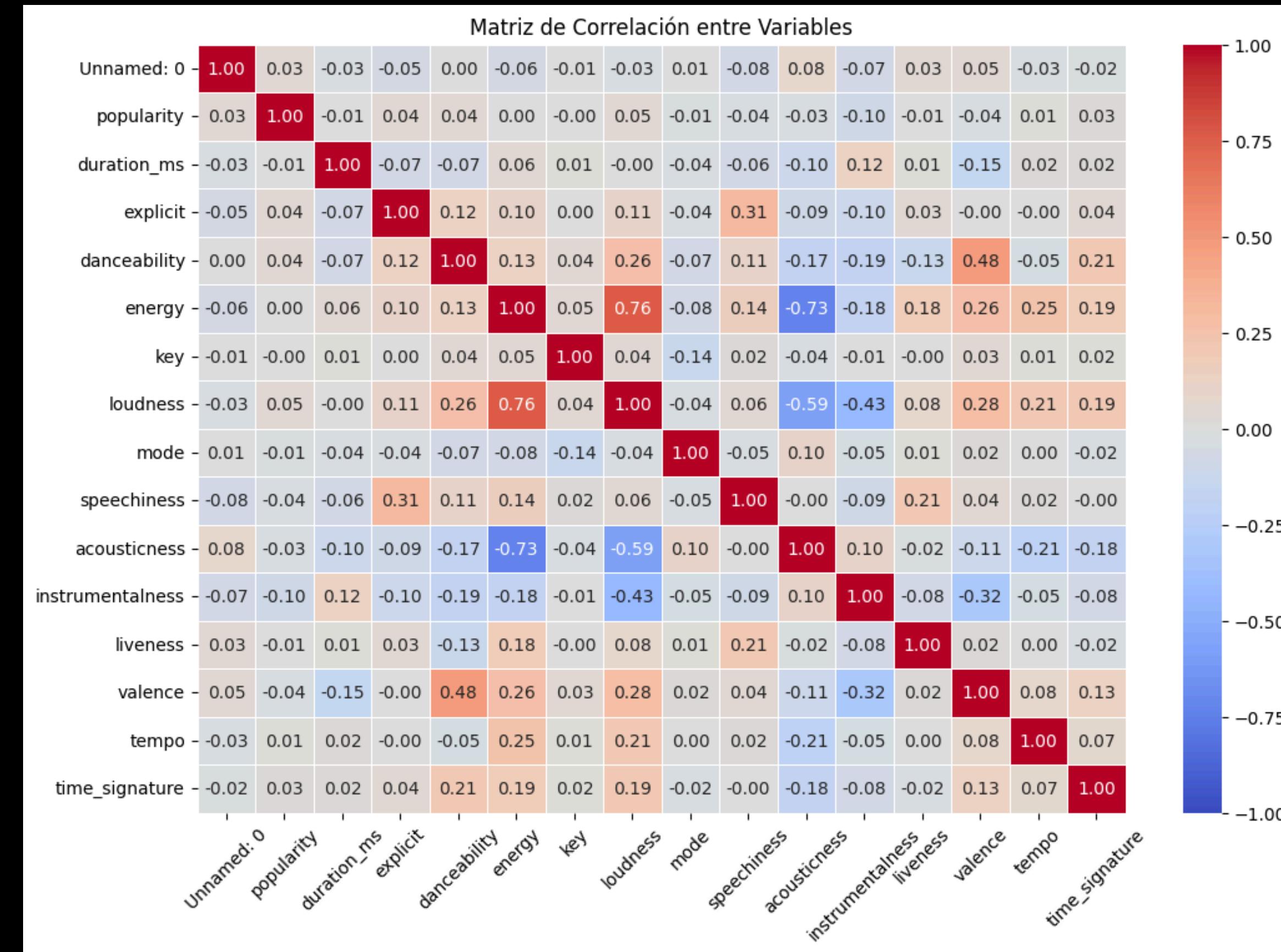


The word "PROGRAMMING" is written in a large, white, sans-serif font, appearing twice in a row. The central letter "O" is replaced by a stylized purple logo consisting of the word "NØIZE" in a bold, rounded font, with a thick purple brushstroke effect running vertically through it.

# El Objetivo

- Predecir el score de popularidad (0-100) de una canción basado en sus características musicales.
- Datos: 114.000+ canciones de Spotify con 20 variables cada una.

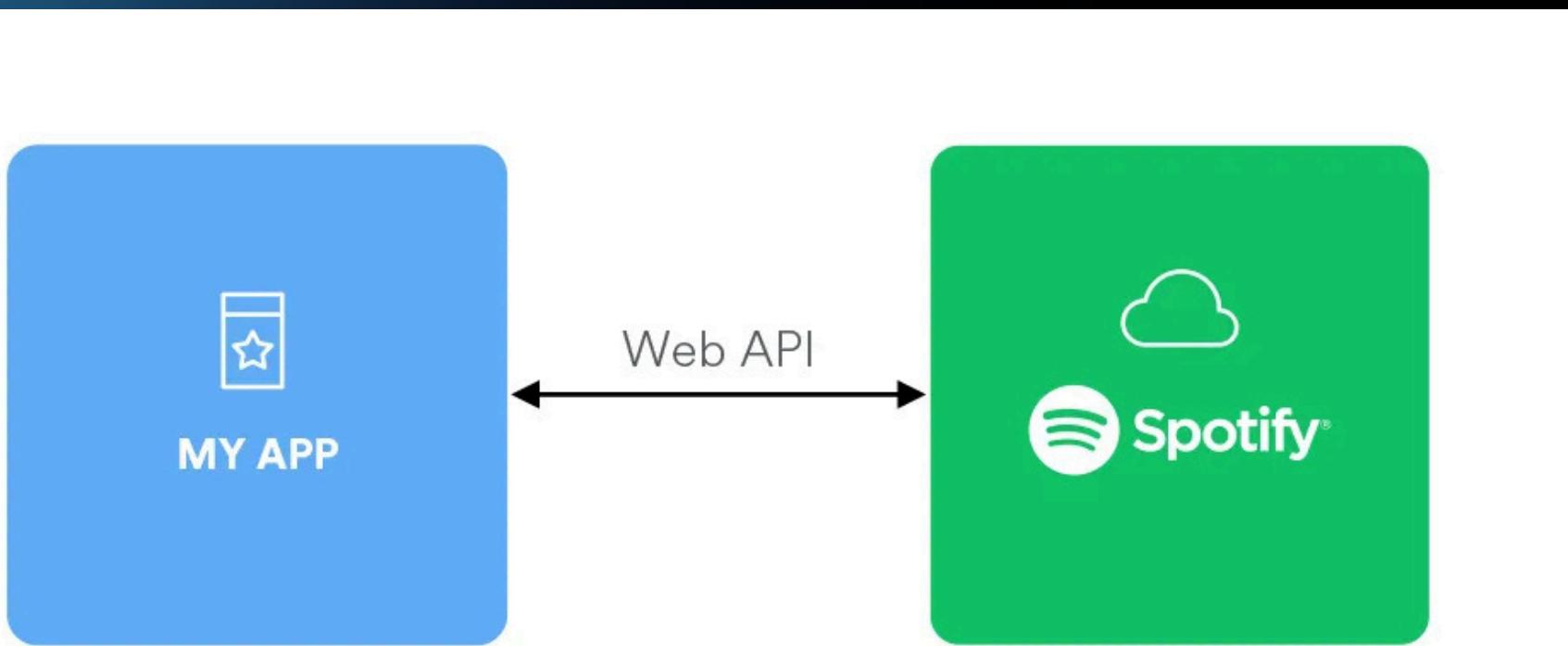






# API

- Terminos y condiciones.
- Tiempo de uso.
- Datos limitados e inconsistencia.





# Limpieza de Datos

- Eliminamos 5% de datos con valores faltantes.
- Quitamos duplicados: 2% del total
- Convertimos 'explícito' de True/False a 1/0
- Aplicación de LabelEncoder() en track\_genre.
- Conversión de unidades de tiempo.



```
17 string sInput;
18 int iLength, iN;
19 double dblTemp;
20 bool again = true;
21 while (again) {
22     iN = -1;
23     again = false;
24     getline(cin, sInput);
25     system("cls");
26     stringstream(sInput) >> dblTemp;
27     iLength = sInput.length();
28     if (iLength < 4) {
29         again = true;
30         continue;
31     } else if (sInput[iLength - 3] != '.') {
32         again = true;
33         continue;
34     } while (++iN < iLength) {
35         if (isdigit(sInput[iN])) {
36             continue;
37         } if (iN == (iLength - 3)) {
```



# Feature Engineering

- Energy & Loudness: Representar mejor la intensidad general de una canción.
- Danceability & Valence: Mejorar la capacidad del modelo para predecir qué tan alegre y bailable es una pista.
- Speechiness / Acousticness: Ayuda a diferenciar canciones con alto contenido vocal respecto a su naturaleza acústica.

# Desempeño de modelos

- Train-test split (80%-20%)
- RMSE y R<sup>2</sup> de los modelos entrenados.
- Hiperparámetros:

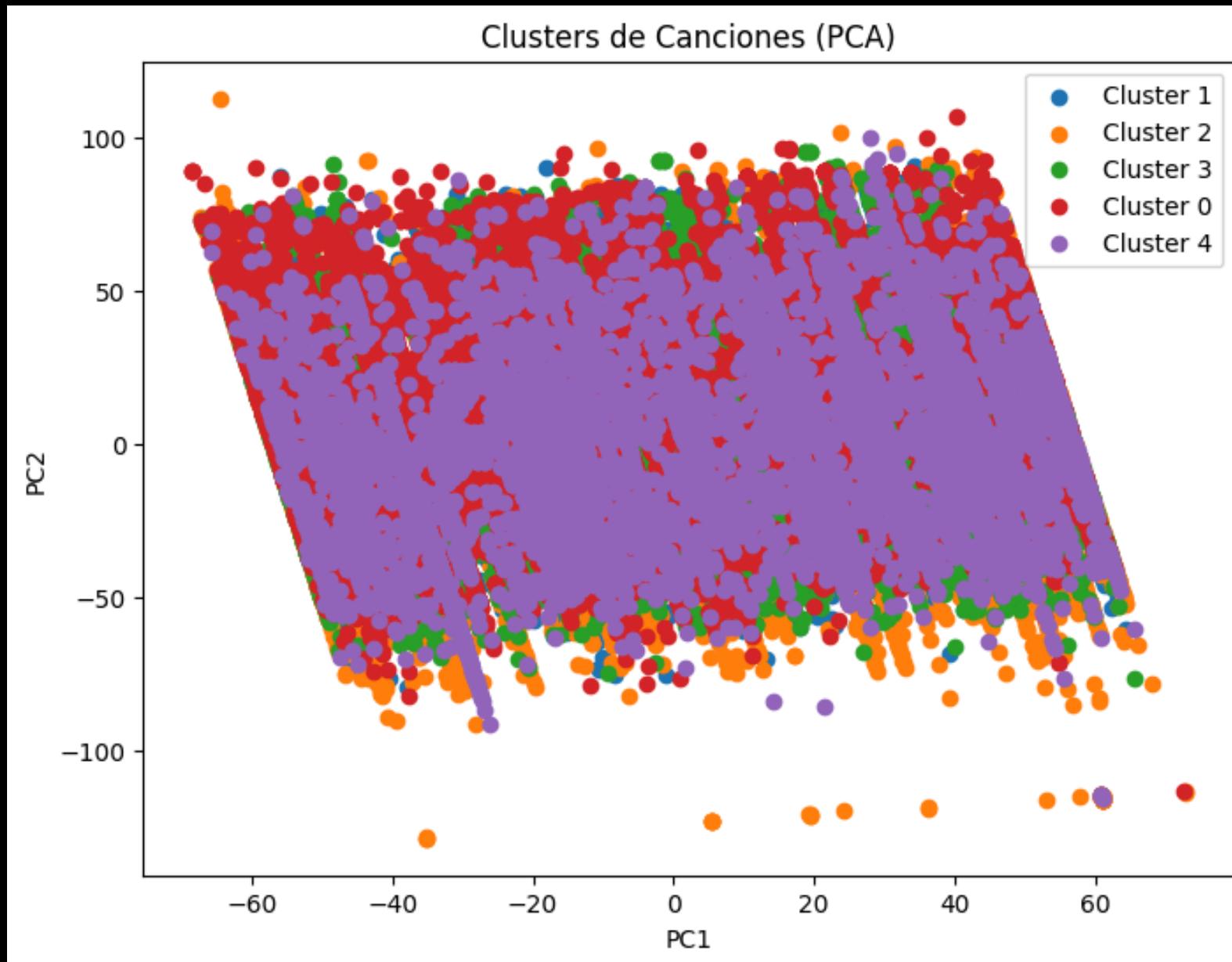
RandomizedSearchCV:

M n\_estimators: 200  
 min\_samples\_split: 5  
 min\_samples\_leaf: 1  
 max\_features: sqrt  
 max\_depth: None

	Modelo	RMSE	R2
1	RandomForest	15.89	0.50
3	XGBoost	17.43	0.39
2	GradientBoosting	20.25	0.18
4	SVR	21.27	0.10
0	LinearRegression	22.09	0.03

	Modelo	RMSE	R2
1	Red Neuronal	21.45	No aplica

# CLUSTERING K-Means

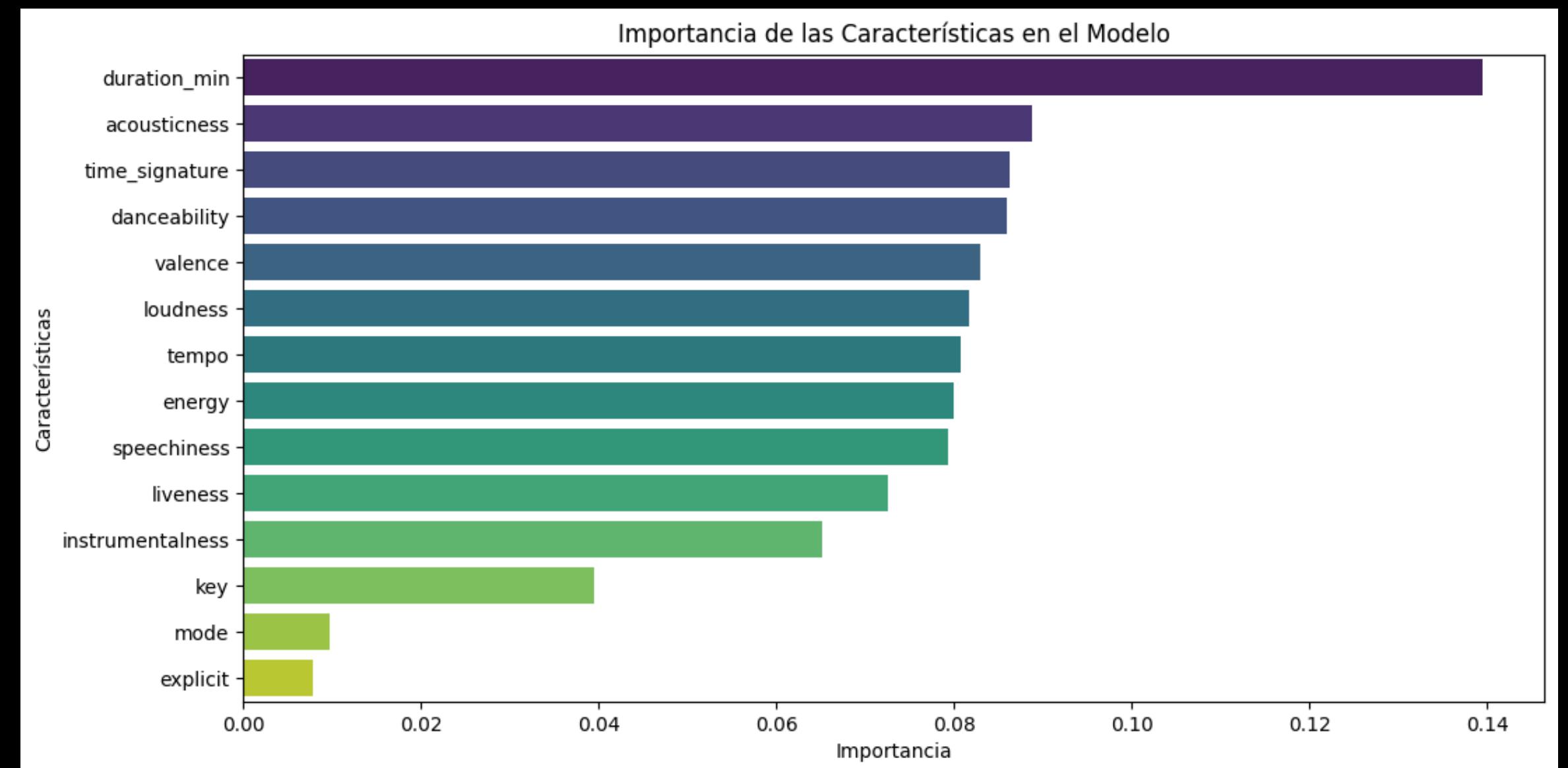


- Descubrir patrones en las canciones sin depender de la variable "popularidad"
- Ayuda a descubrir patrones en las canciones sin depender de la variable "popularidad", para identificar posibles tendencias.
- Usar los clusters como una nueva variable en el modelo supervisado.

	explicit	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_min	track_genre_encoded	
cluster	0	0.034335	0.471658	0.801470	5.372059	-6.386492	0.589569	0.076808	0.066995	0.315755	0.190681	0.285241	137.674506	4.673727	46.738613
1	0.294187	0.673177	0.709524	5.956892	-6.465842	0.156839	0.105346	0.206253	0.042538	0.171353	0.588808	119.911214	3.525880	57.272197	
2	0.006986	0.448280	0.261240	4.931689	-15.108377	0.705096	0.050514	0.792294	0.335445	0.158203	0.284471	109.115540	3.567790	59.709128	
3	0.000000	0.639592	0.690123	5.001436	-6.764005	1.000000	0.070012	0.265101	0.034484	0.172063	0.638875	121.255099	3.502682	60.255451	
4	0.121684	0.521246	0.747581	5.272959	-7.675157	0.682781	0.195080	0.357983	0.064415	0.755702	0.492959	120.785466	3.910205	59.887117	

# ¿Por qué Random Forest?

- No lineal: Captura relaciones complejas o en otras palabras no lineales (como interacciones entre 'danceability' y 'energy').
- Robusto: Outliers en loudness no arruinan el modelo.
- Normalización: No requiere normalización, lo que simplifica el preprocessamiento.

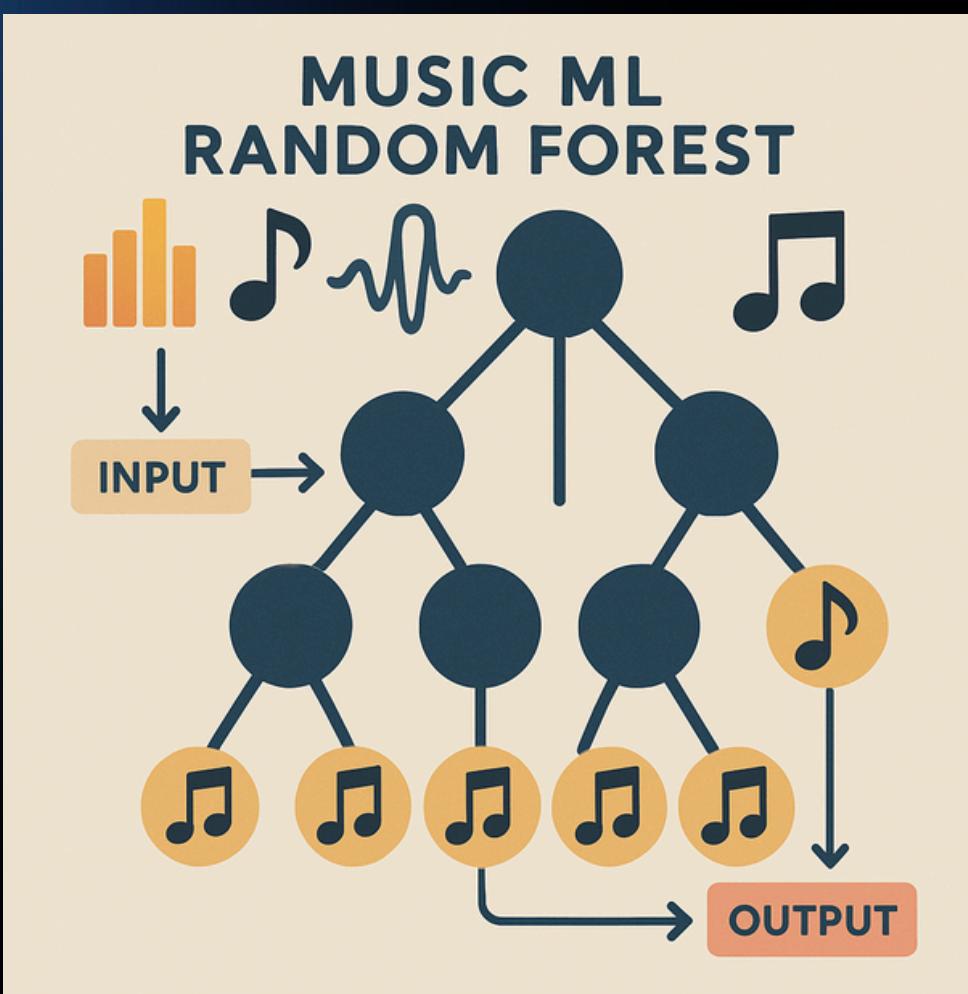




# Optimización



- Añadir más datos o features relevantes (ej: letras, datos de redes sociales).
- Usar una nueva variable Cluster.
- Arquitecturas más complejas en futuras versiones de redes neuronales.



Innovation

Connection

Impact

Revolution

THANK YOU

