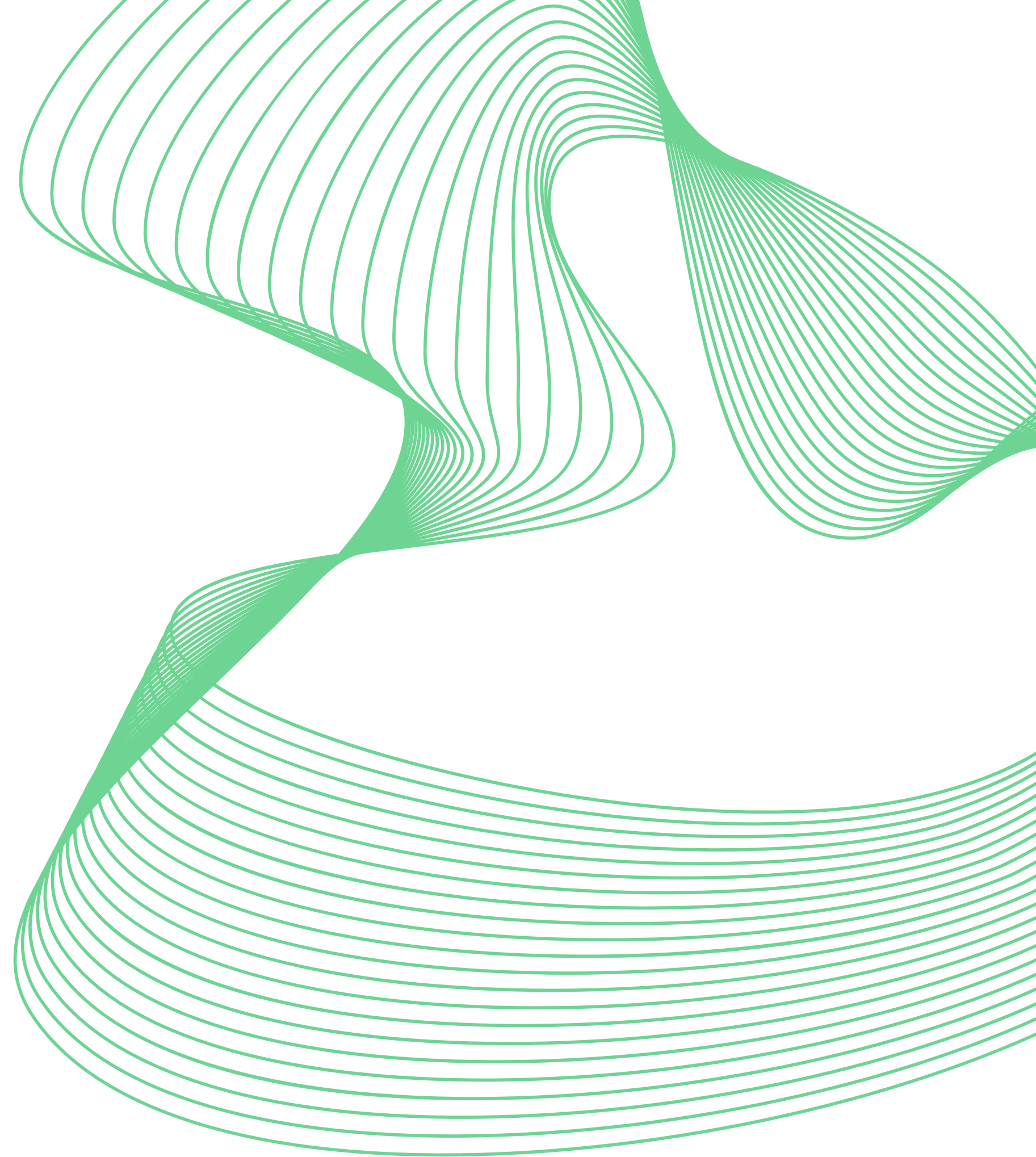


Propuesta MACHINE LEARNING



Introducción

- Predecir la “danceability” de las canciones para abordar necesidad de negocio de Spotify.
- Análisis de datos
- Modelo ML



Análisis de datos

Dataset de Música → **Kaggle**

No valores nulos + Estandarizados
Sí NaN

50.000 filas → Canciones

17 columnas → Variables: popularidad, volumen, energía..



Volumen (dB) 

Valencia 

Energía 

Habla/Letra 

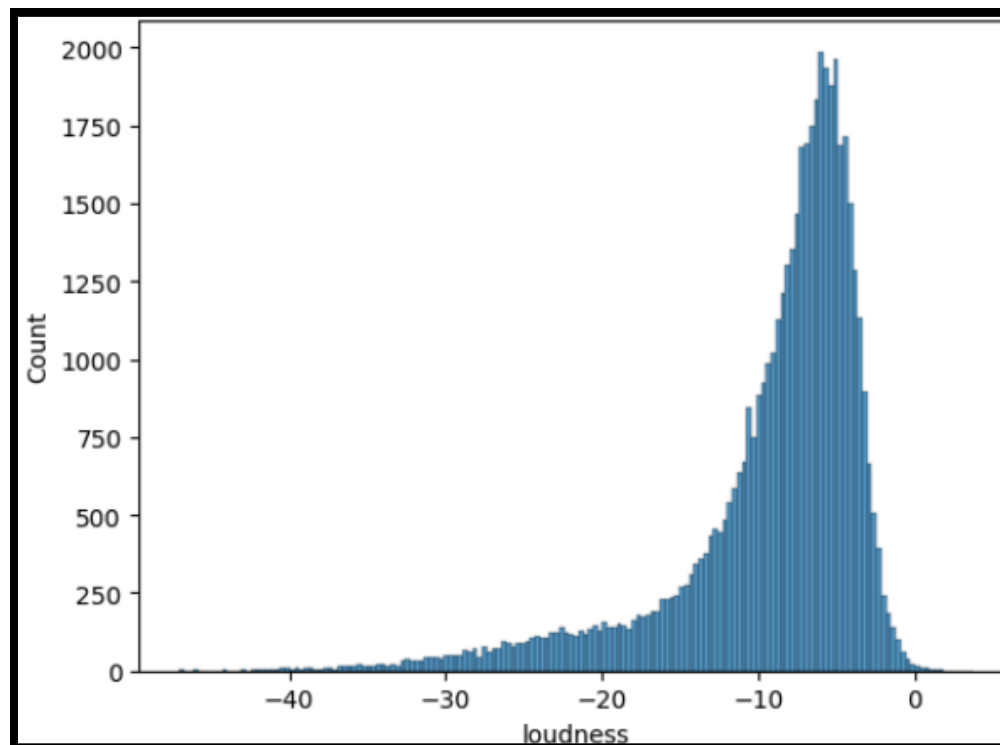
Acústica 

Instrumentalidad 

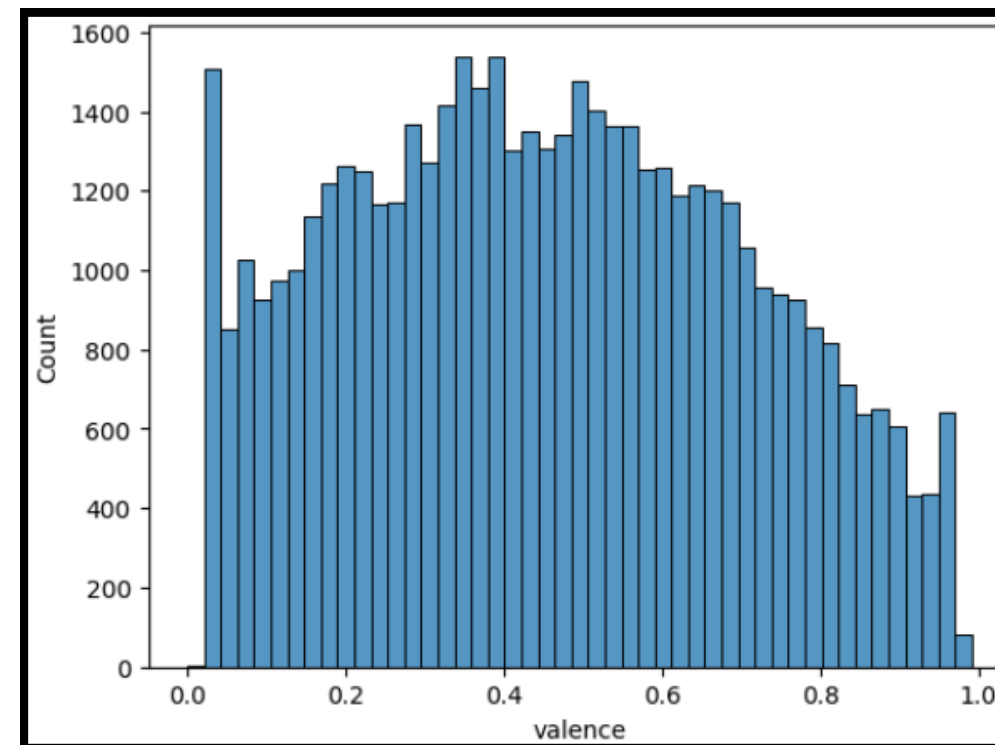
Popularidad 

Análisis de datos

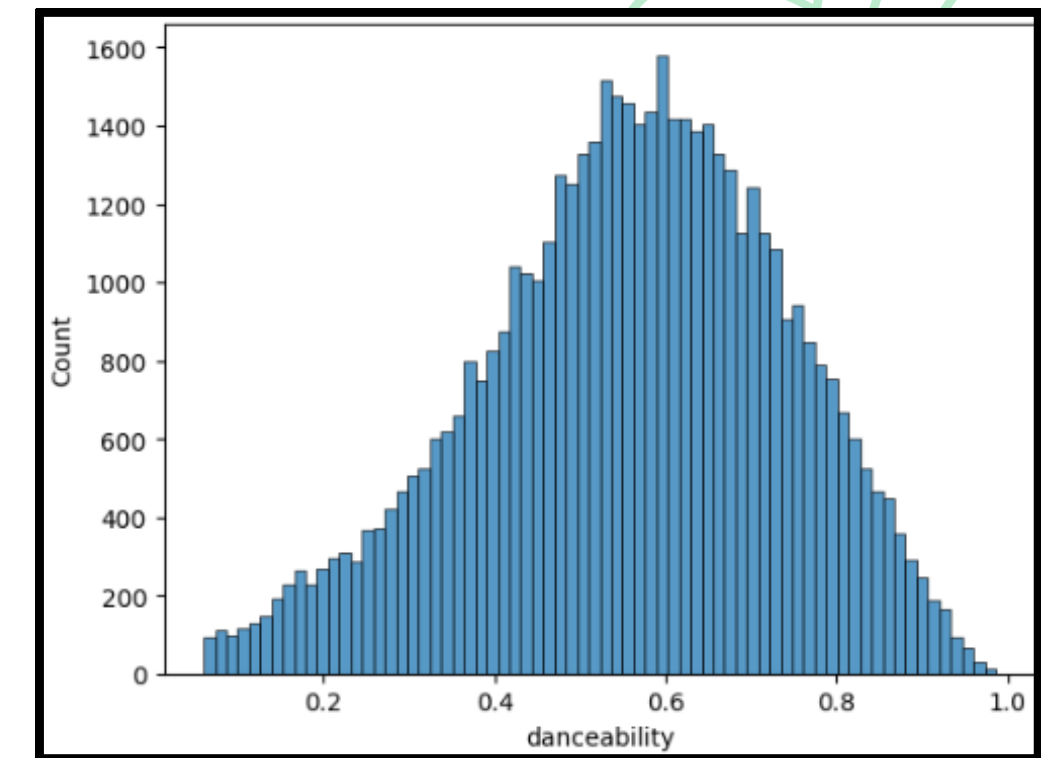
Volumen (dB)



Valencia

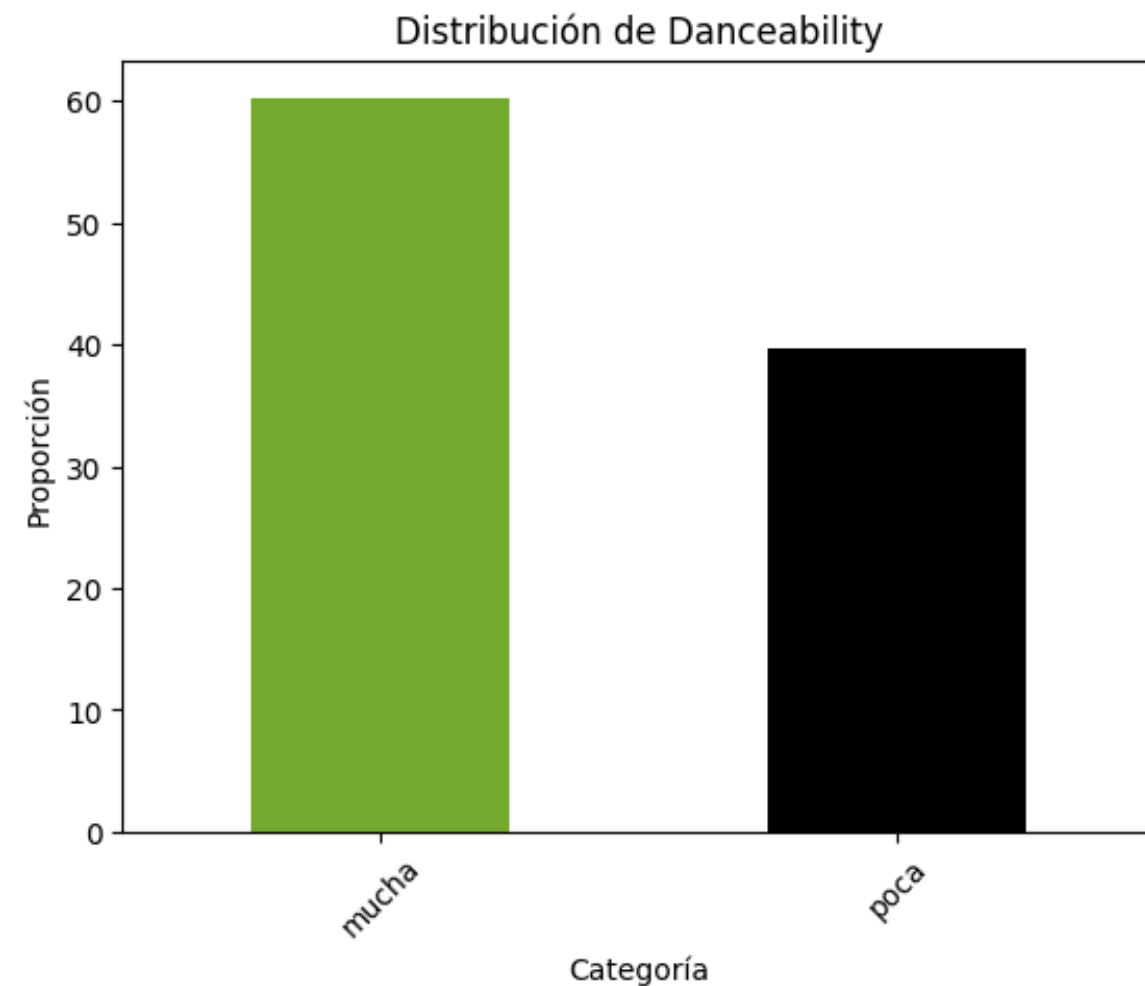


Danzabilidad



Preprocesamiento de datos

Distribución del target



Codificación del target

Mucha 'danceability' 60%
Poca 'danceability' 40%



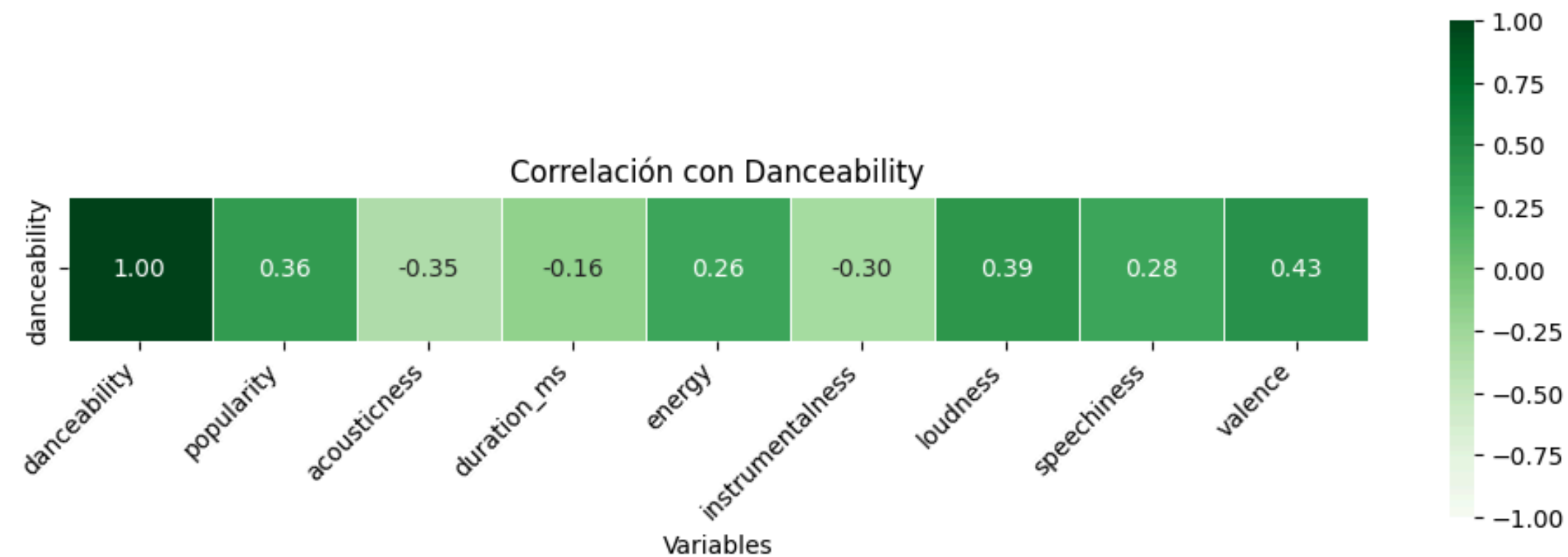
1
0



Problema de clasificación

Preprocesamiento de datos

Correlaciones con el target



Positivamente correladas

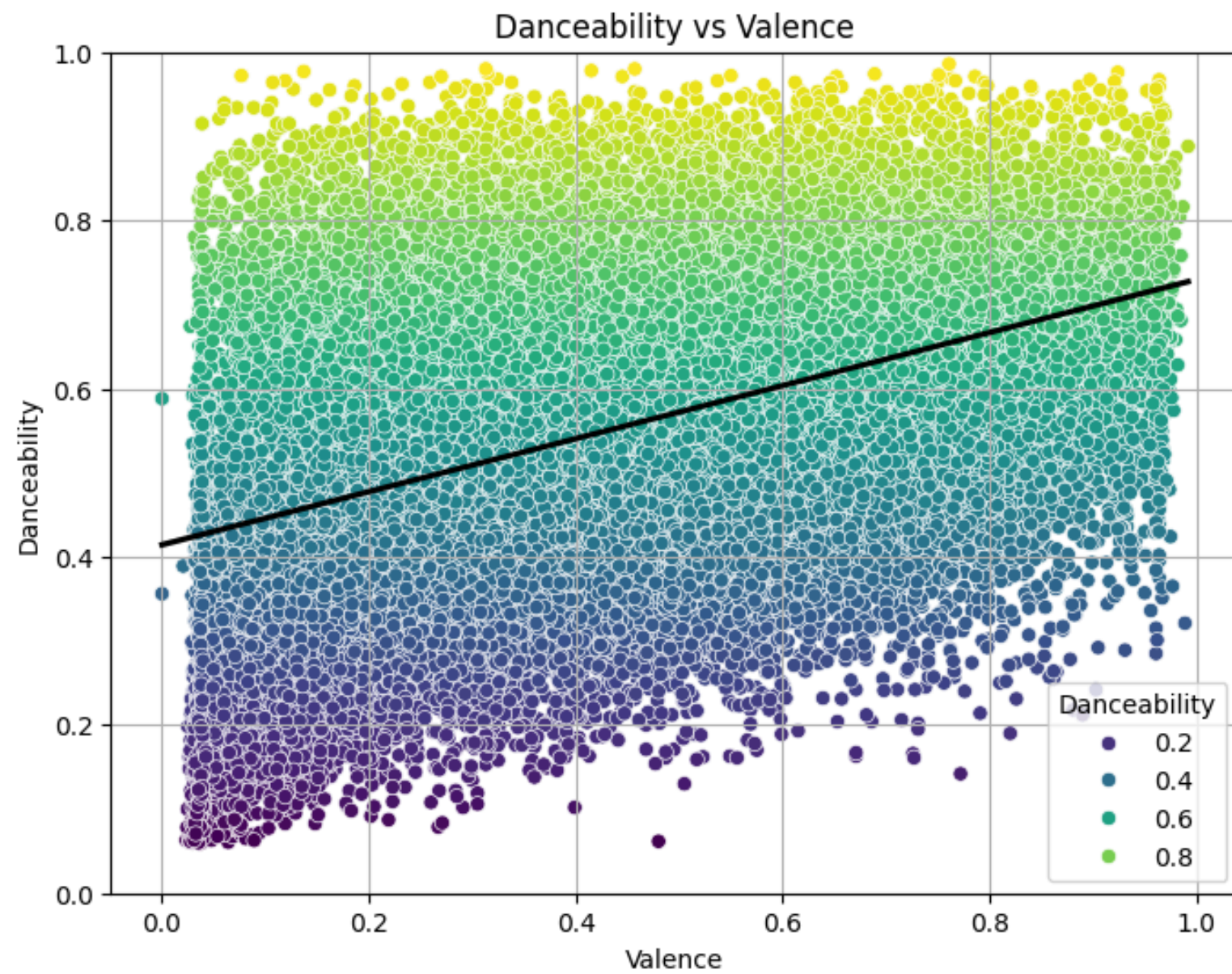
- **Popularidad** (0.36)
- **Energía** (0.26)
- **Volumen** (0.39)
- **Hablabilidad** (0.28)
- **Valencia** (0.43)

Negativamente correladas

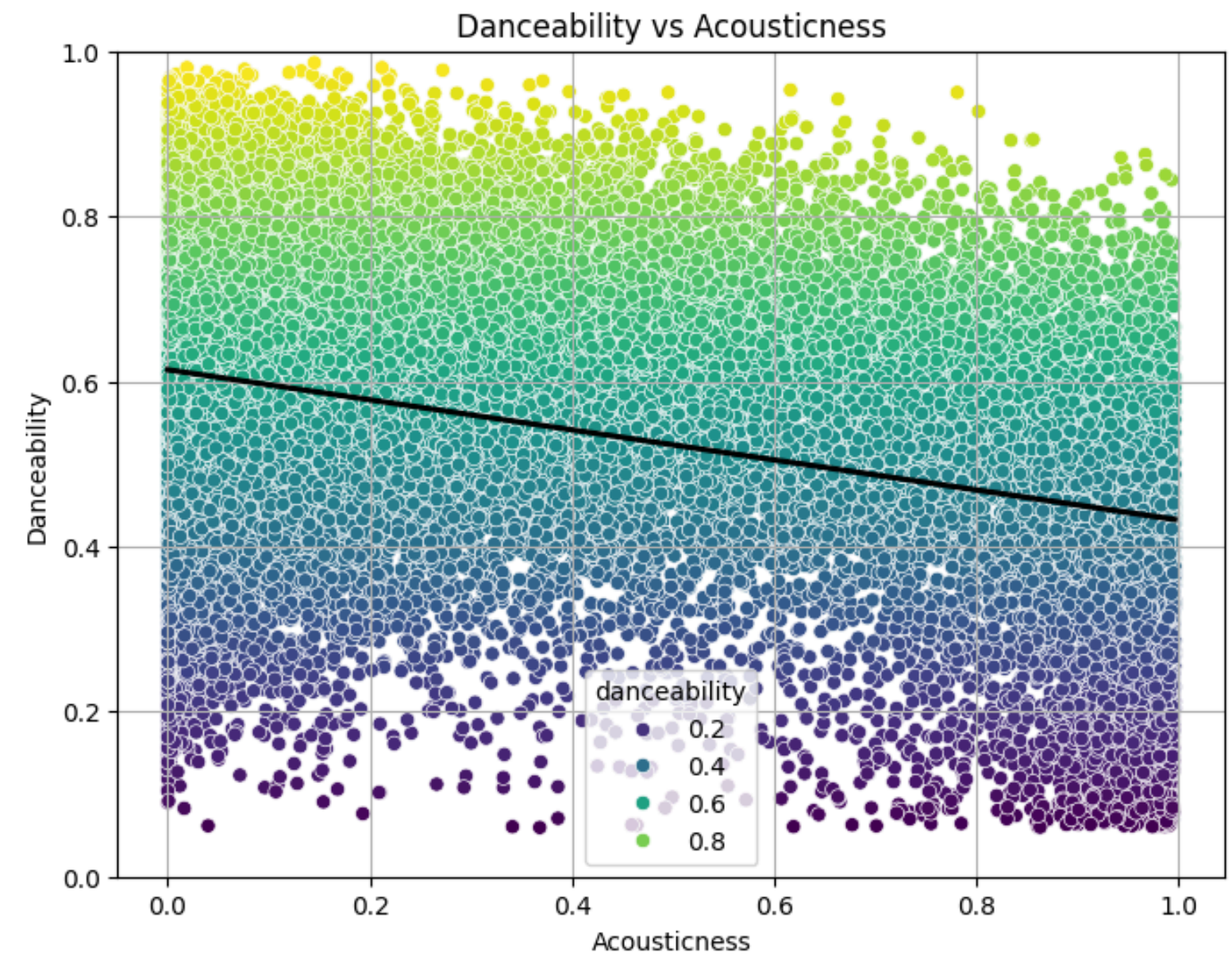
- **Acústica** (-0.35)
- **Instrumentalidad** (-0.3)
- **Duración** (-0.16)

Correlaciones

Valencia vs Danzabilidad



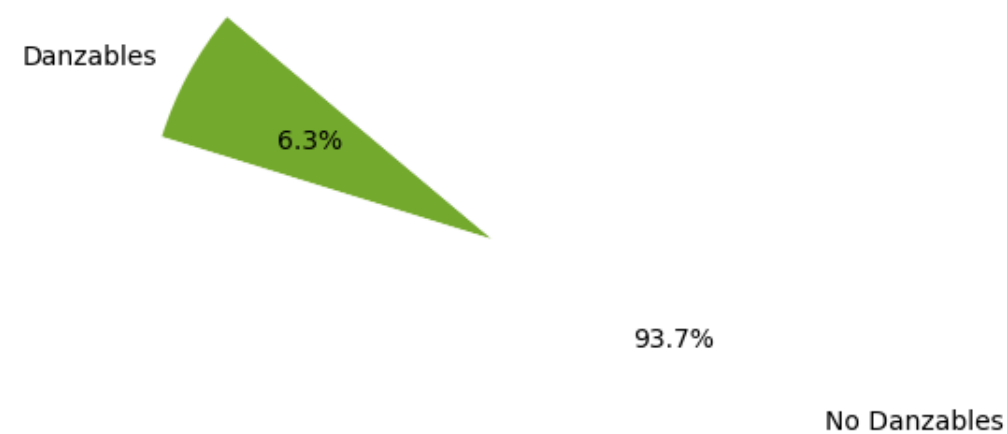
Acústica vs Danzabilidad



Entendiendo correlaciones

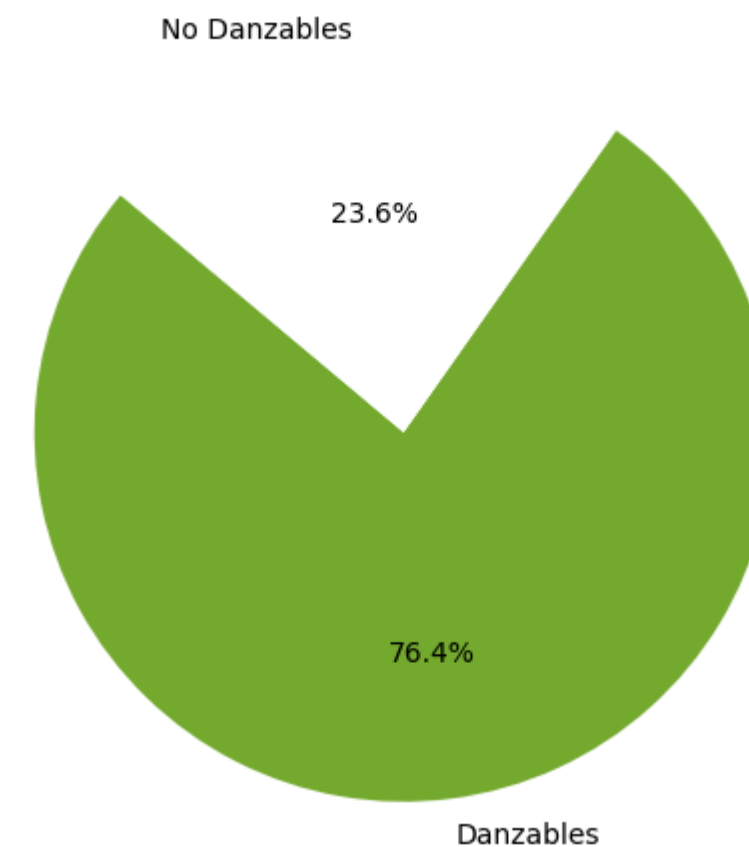
Música clásica vs. danzabilidad

% de canciones del género "Clásica" que son danzables



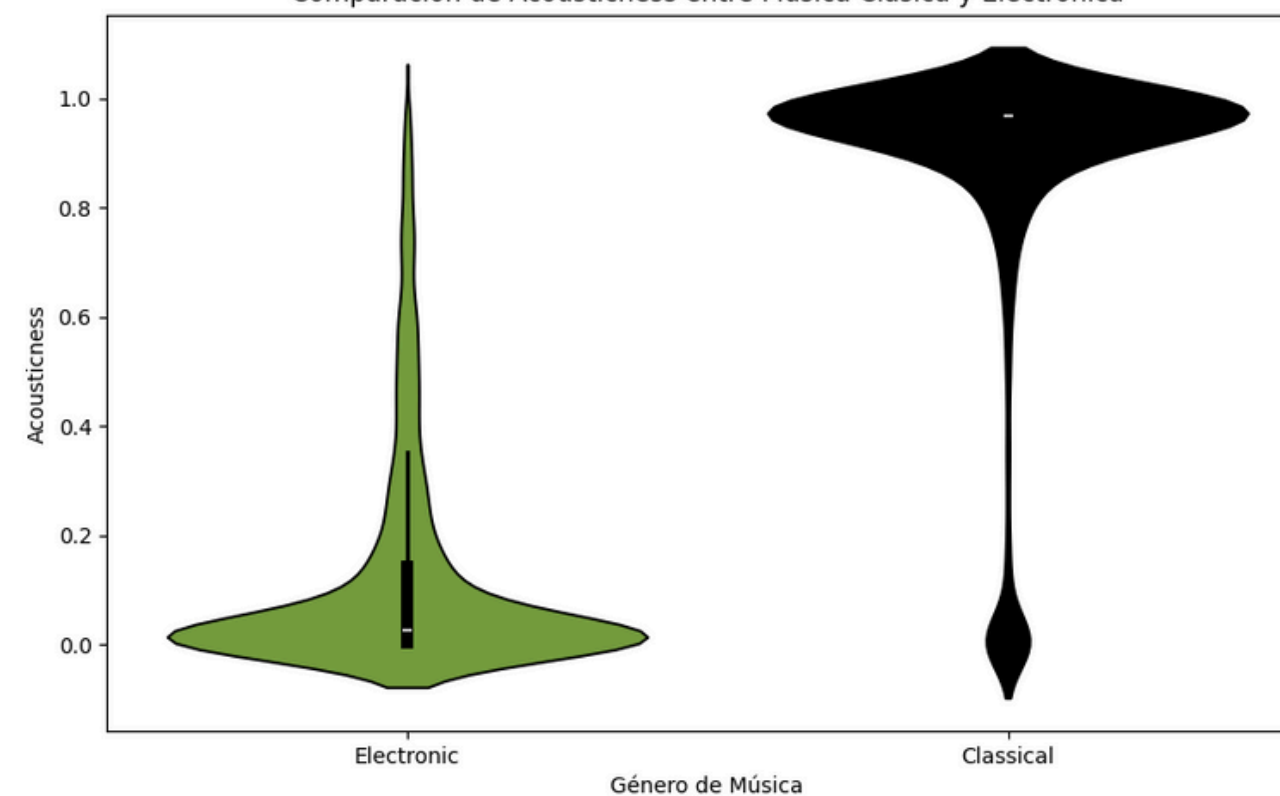
Música electrónica vs. danzabilidad

% canciones del género "Electrónica" que son danzables



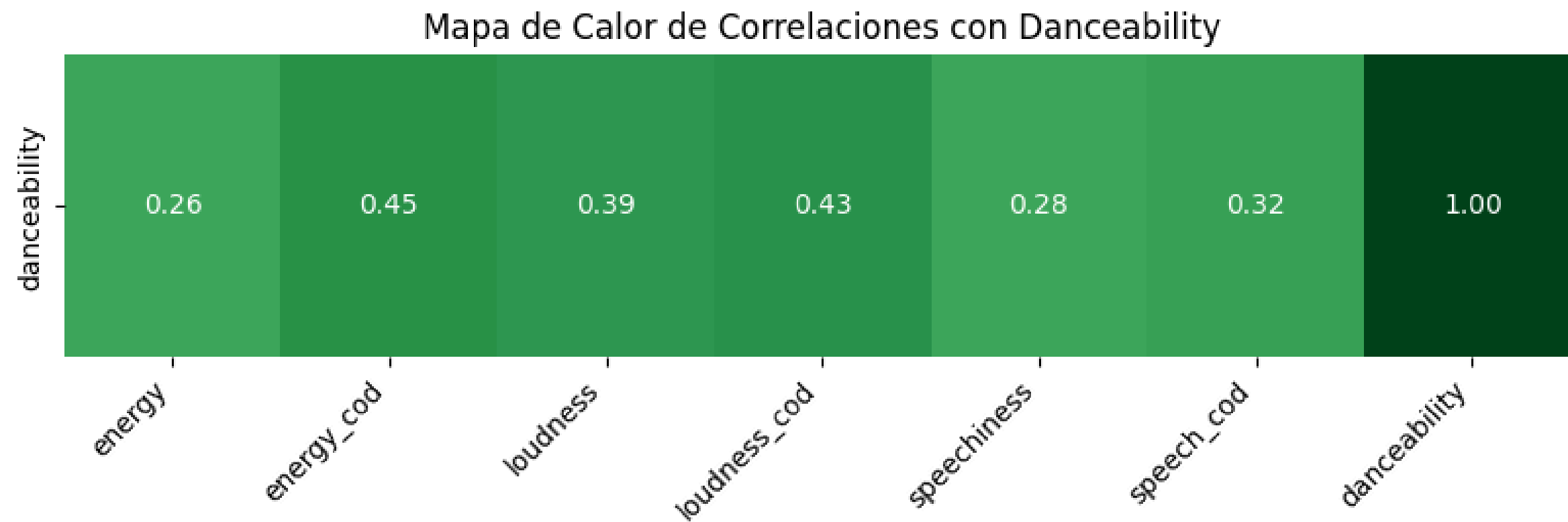
Acústica por género musical

Comparación de Acousticness entre Música Clásica y Electrónica



Aumentando correlaciones

1. Dividir la variable en rangos y asignar números
2. Agrupar la variable por la media de "danceability"
3. Reajustar los valores si fuera necesario



- Energía 0.26 – **0.45**
- Volumen 0.39 – **0.43**
- Habla/Letra 0.28 – **0.32**

Modelado

Variables predictoras útiles

- Popularidad
- **Energía**
- **Volumen**
- **Letra**
- Valencia
- Acústica
- Instrumentalidad
- Tempo
- Género

Target

- Danzabilidad

Resultados sobre TEST

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
lightgbm	Light Gradient Boosting Machine	0.8051	0.8778	0.8779	0.8134	0.8444	0.5846	0.5877
rf	Random Forest Classifier	0.8045	0.8746	0.8785	0.8123	0.8441	0.5831	0.5864
xgboost	Extreme Gradient Boosting	0.8034	0.8736	0.8724	0.8146	0.8425	0.5819	0.5845
et	Extra Trees Classifier	0.8001	0.8702	0.8725	0.8102	0.8402	0.5742	0.5771
gbc	Gradient Boosting Classifier	0.7997	0.8711	0.8853	0.8026	0.8419	0.5704	0.5757
ada	Ada Boost Classifier	0.7879	0.8577	0.8698	0.7968	0.8317	0.5464	0.5504
qda	Quadratic Discriminant Analysis	0.7362	0.7972	0.8856	0.7325	0.8018	0.4179	0.4351
lr	Logistic Regression	0.7323	0.7919	0.8534	0.7414	0.7934	0.4184	0.4270
knn	K Neighbors Classifier	0.7323	0.7754	0.8321	0.7506	0.7892	0.4249	0.4293