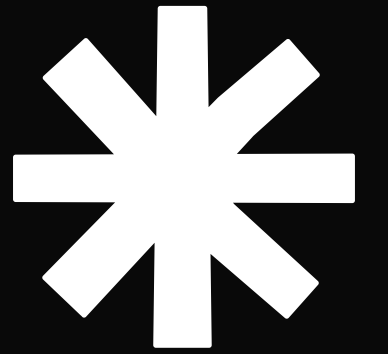


Construcción y Evaluación de un Modelo de Clasificación Musical



Lucas Herranz

CONTEXTO I



Problemática

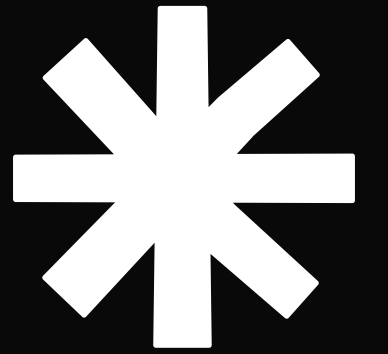
La clasificación de géneros musicales, a partir de archivos de audio.

Datos

- Tipos:
 - Archivos .wav organizados en carpetas por género.
 - Un CSV con id's, path, nombre.
- Tamaño: ~50GB de datos de audio almacenados en Google Drive



CONTEXTO II

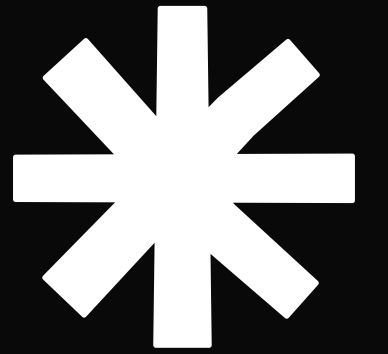


Técnicas

- La librería Librosa de python para extracción de características de audio.
- Modelos:
 - Modelos Clásicos (XGBoost, CatBoost, LightGBM, Random Forest): Comparación de enfoques tradicionales vs. deep learning.
 - Redes Neuronales : Para aprendizaje sobre espectrogramas.
- Evaluación:
 - Métricas: Accuracy, F1-score, Matriz de Confusión.
 - Cross-validation para garantizar generalización.



CONTEXTO III



Variables

Predictoras:

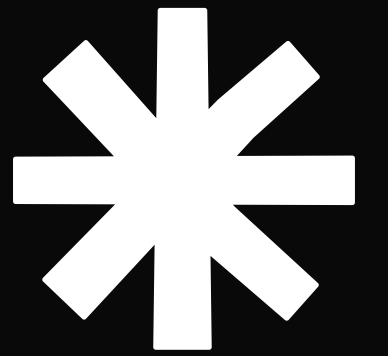
- MFCCs (Mel-Frequency Cepstral Coefficients): Capturan características del timbre y son clave para la clasificación.
- Chroma Features: Relacionadas con la tonalidad y armonía.
- Spectral Contrast: Diferencias en energía entre bandas de frecuencia, útil para distinguir géneros.
- Zero-Crossing Rate: Indica la cantidad de cambios de signo en la onda, relevante para distinguir sonidos percusivos.
- Tempo/BPM: Velocidad de la música, útil para diferenciar géneros rítmicamente distintos.

Target:

- Género musical



EDA



19.896
Pistas de audio

19
Géneros

20
MFCC

12
Chroma Features

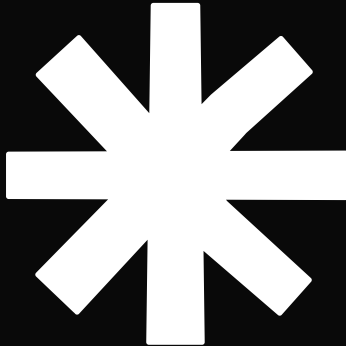
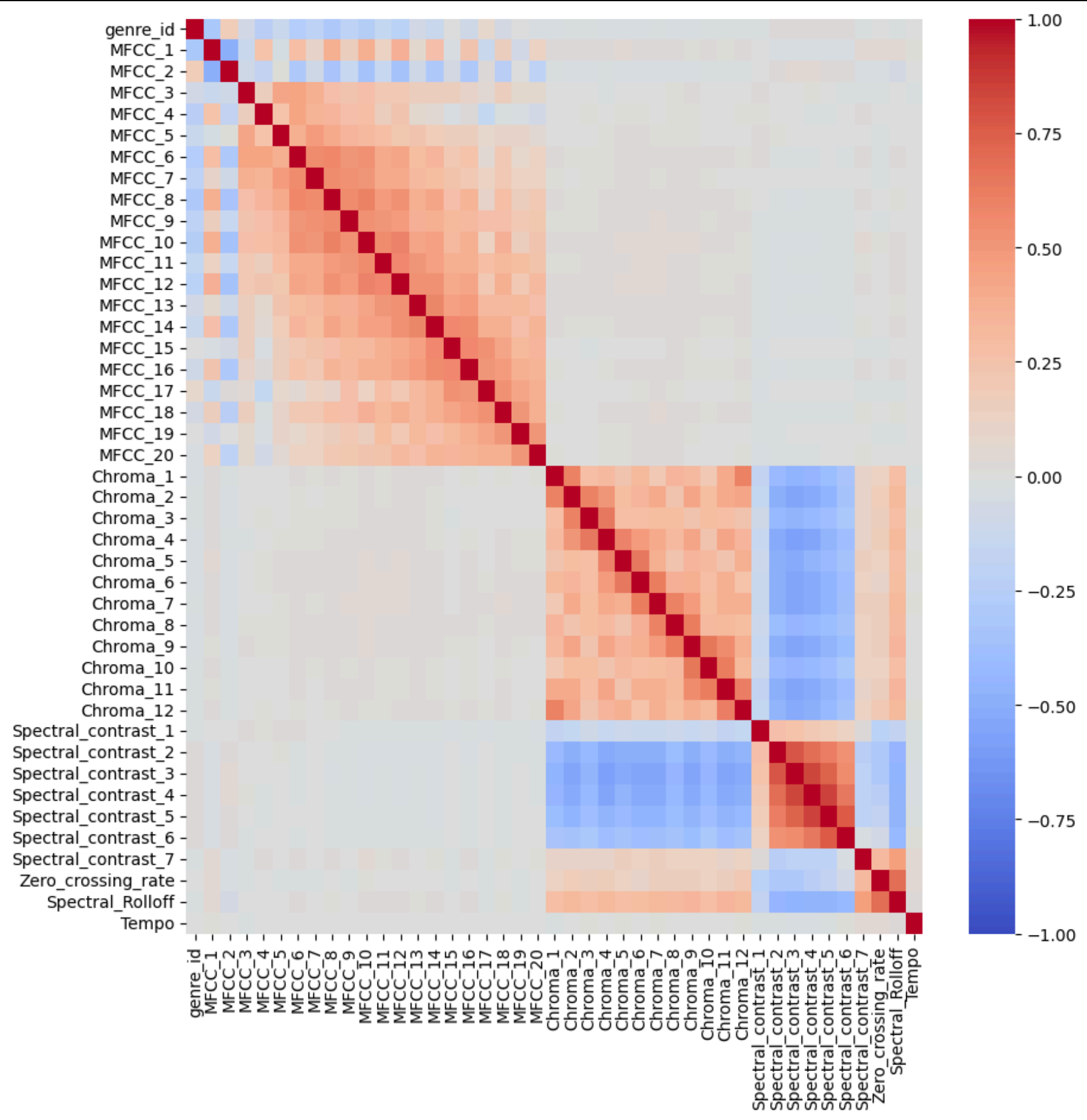
7
Spectral Contrast

1
Zero-Crossing Rate

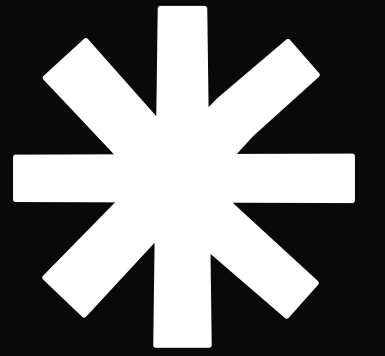
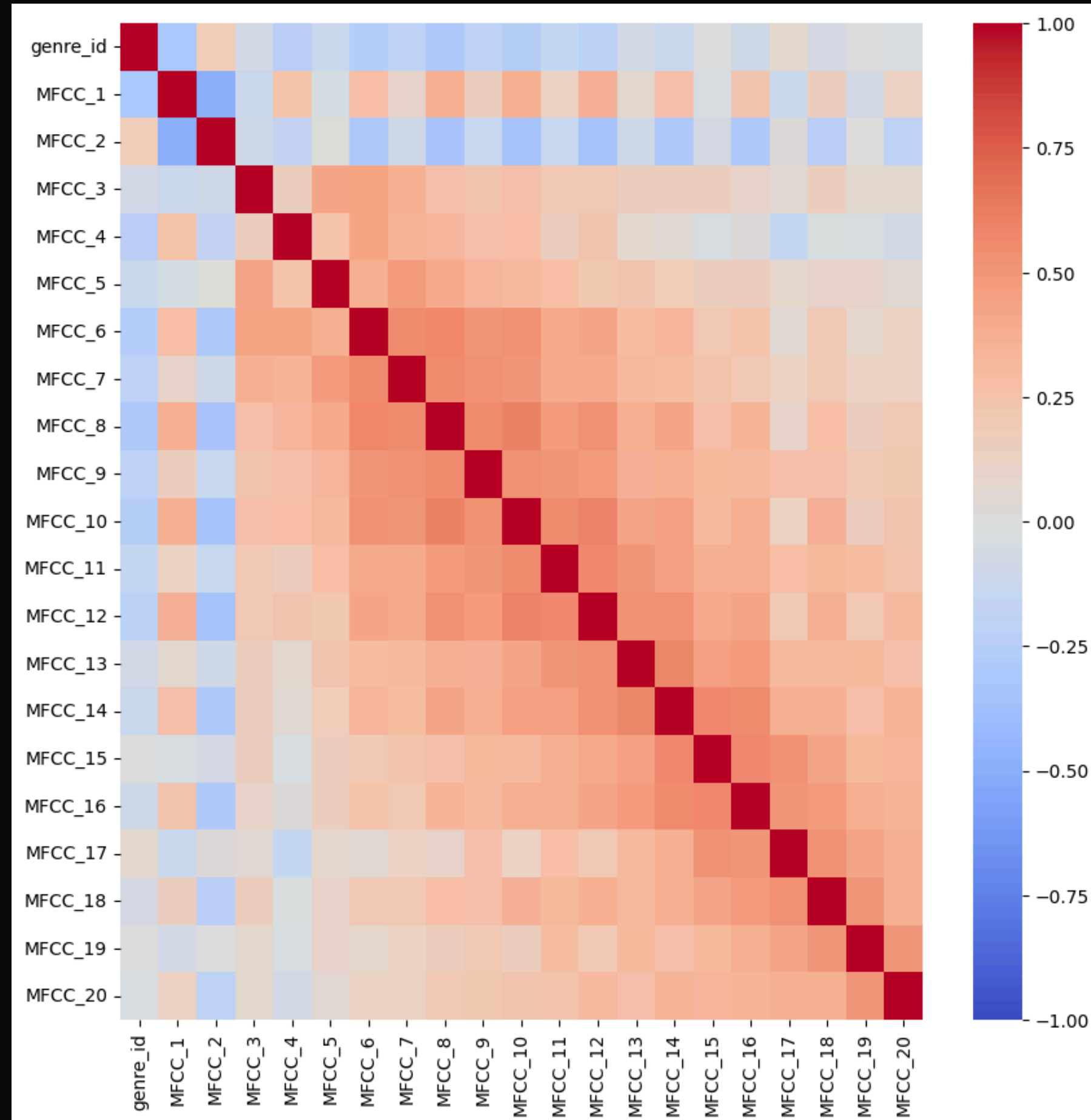
1
Tempo



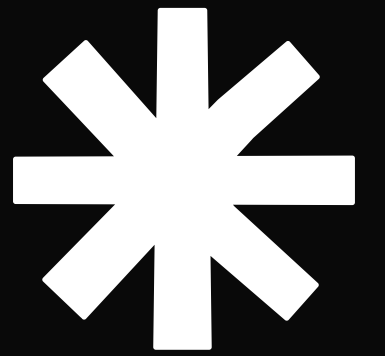
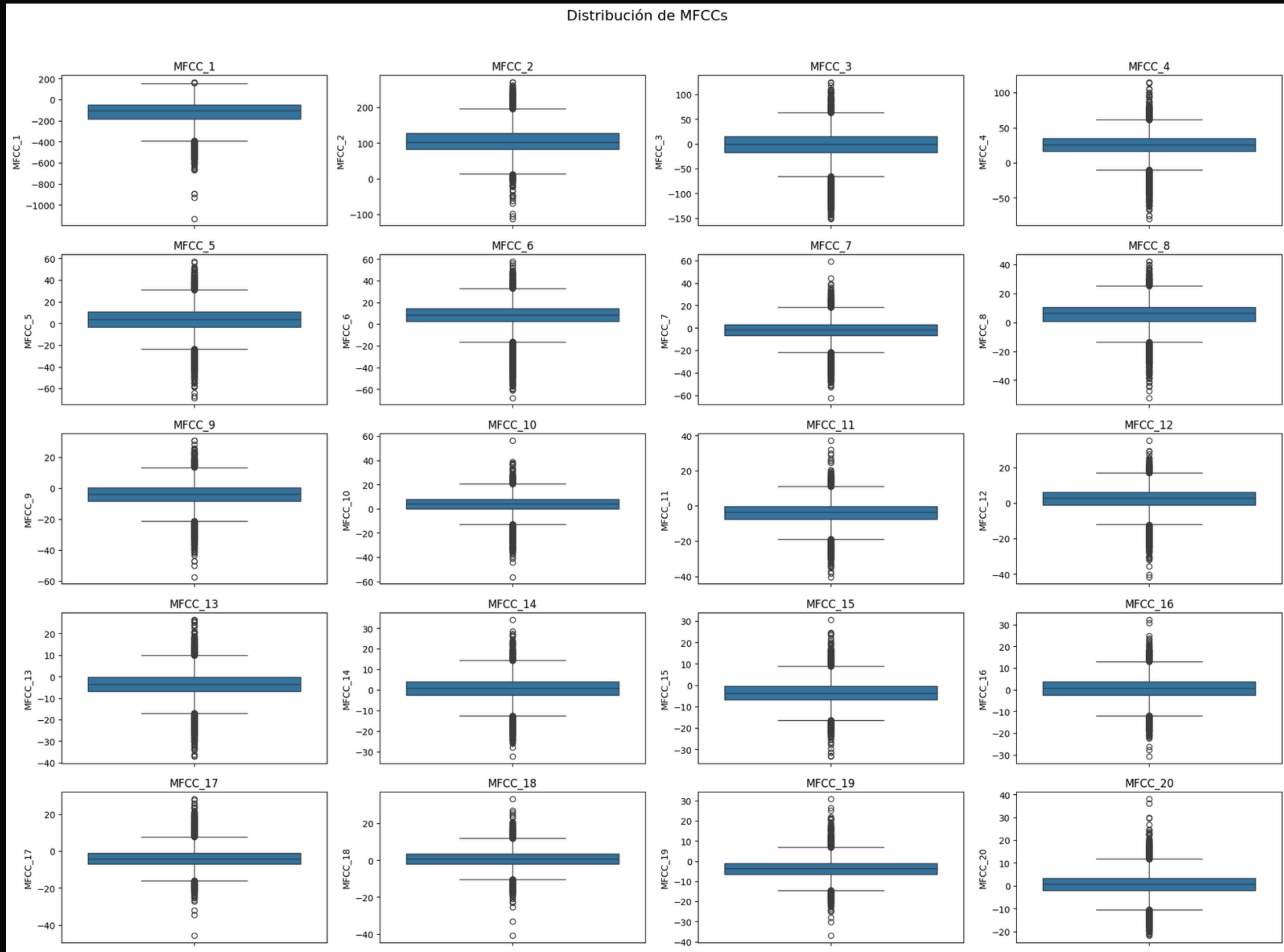
EDA



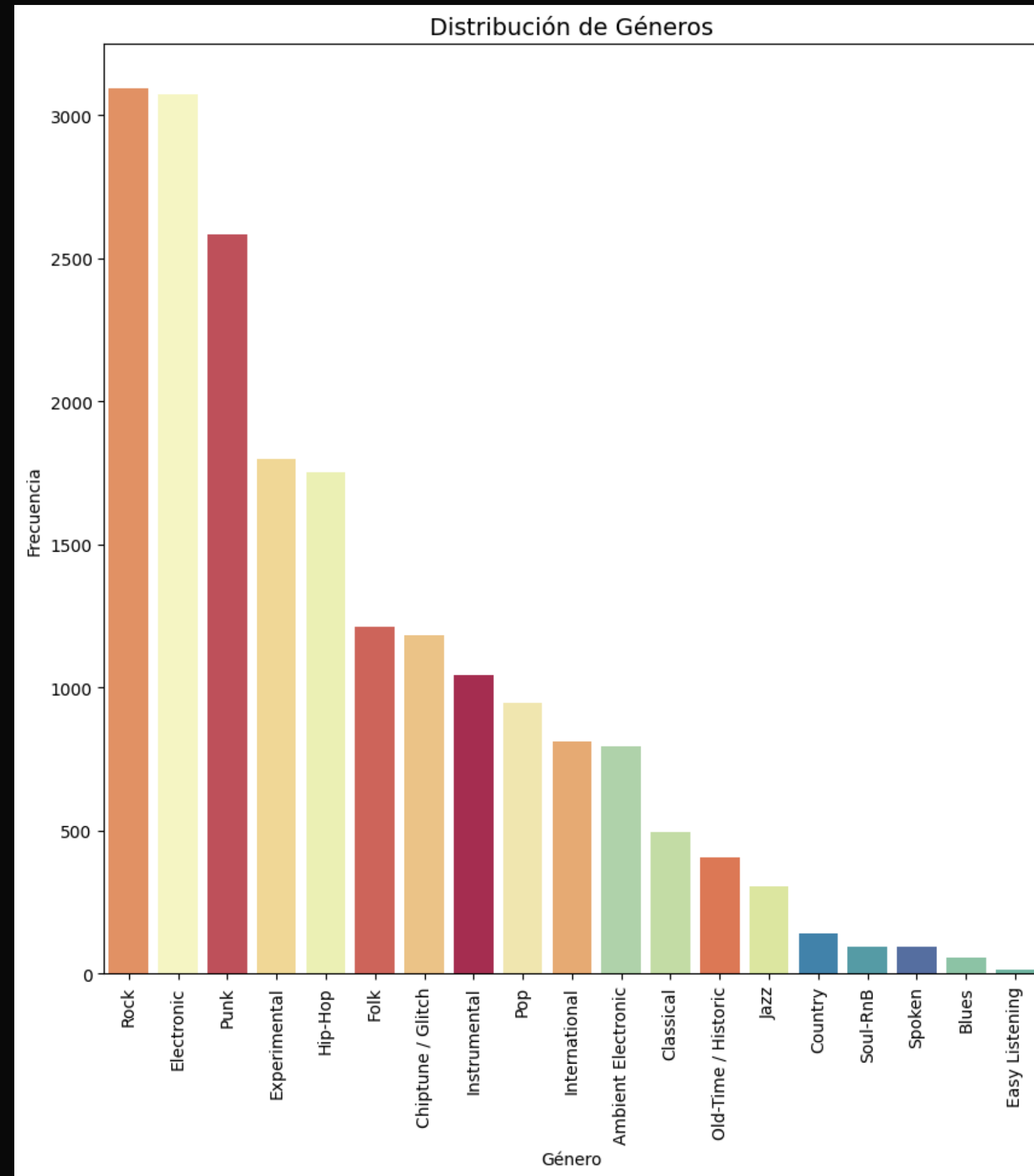
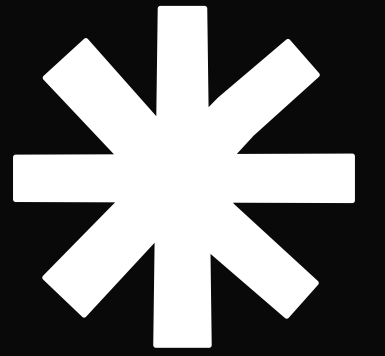
EDA



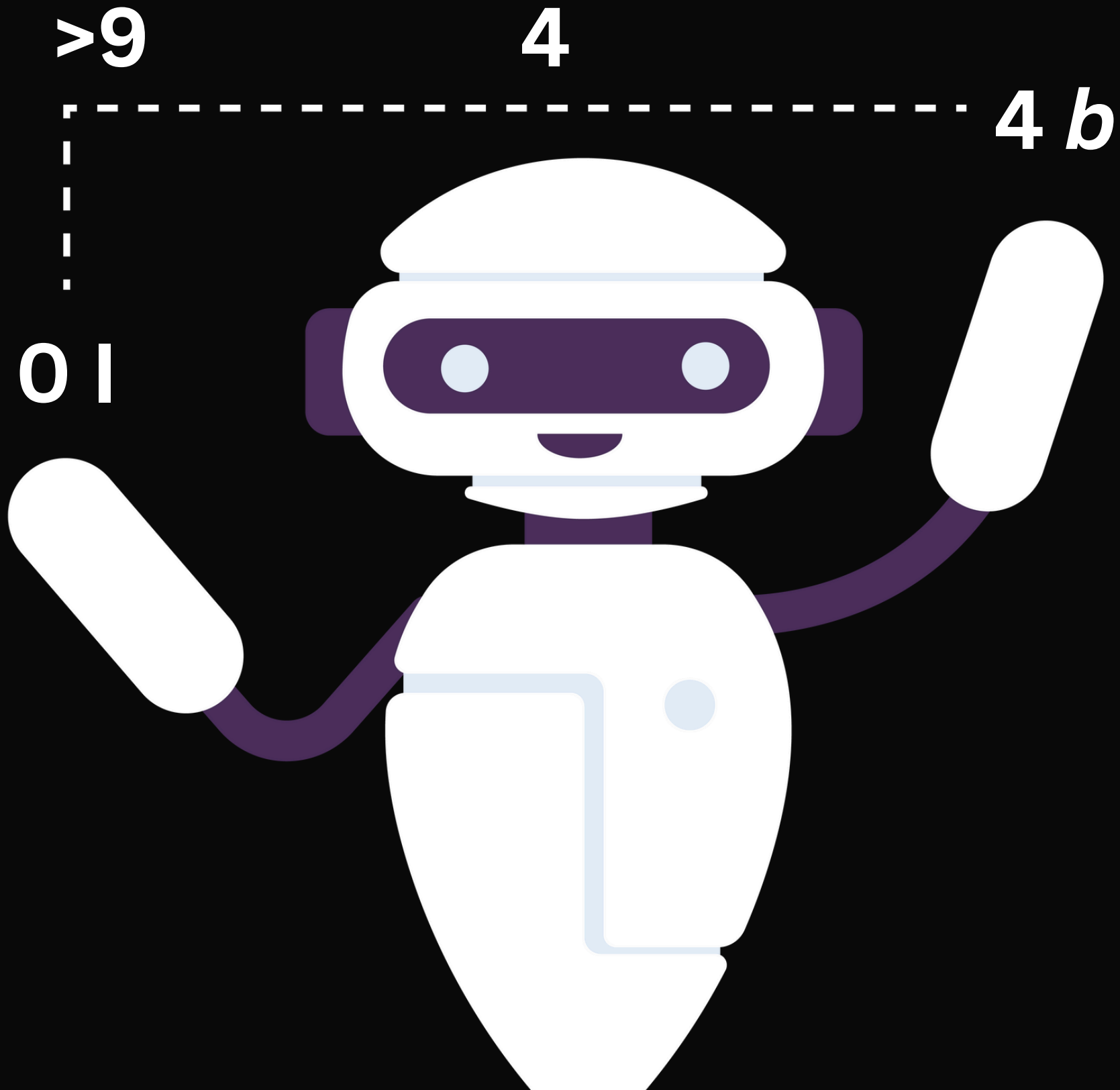
EDA



EDA



APROXIMACIONES



APROXIMACIÓN 0

Problema binario: Música digital vs Música estudio

APROXIMACIÓN 1

Nueve géneros con mayor número de muestras

APROXIMACIÓN 2

Agrupación de los géneros en 4 super géneros

APROXIMACIÓN 3

Balanceo de target en los 4 super géneros

melodIA

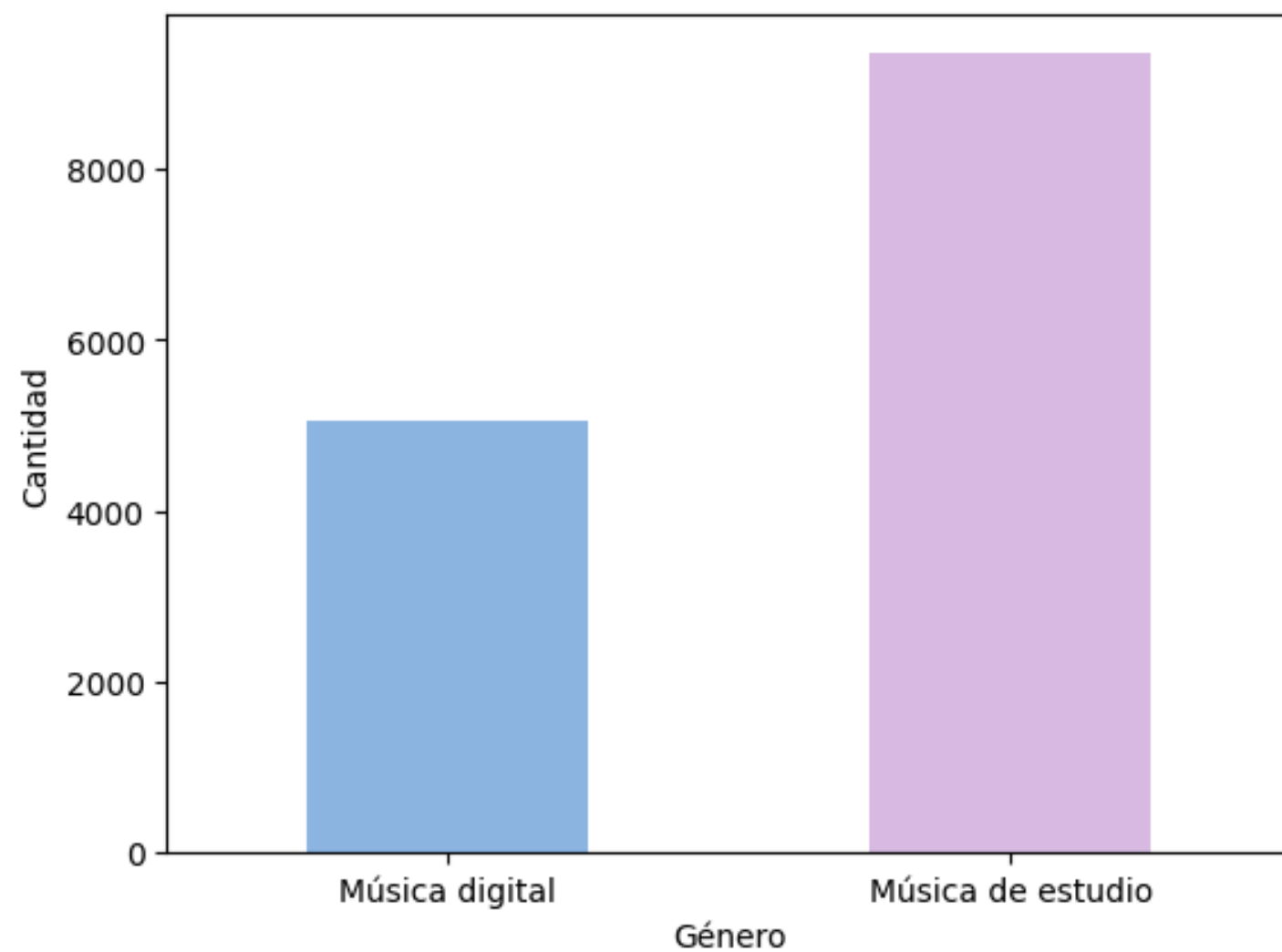


APROXIMACIÓN 0 - CatBoost

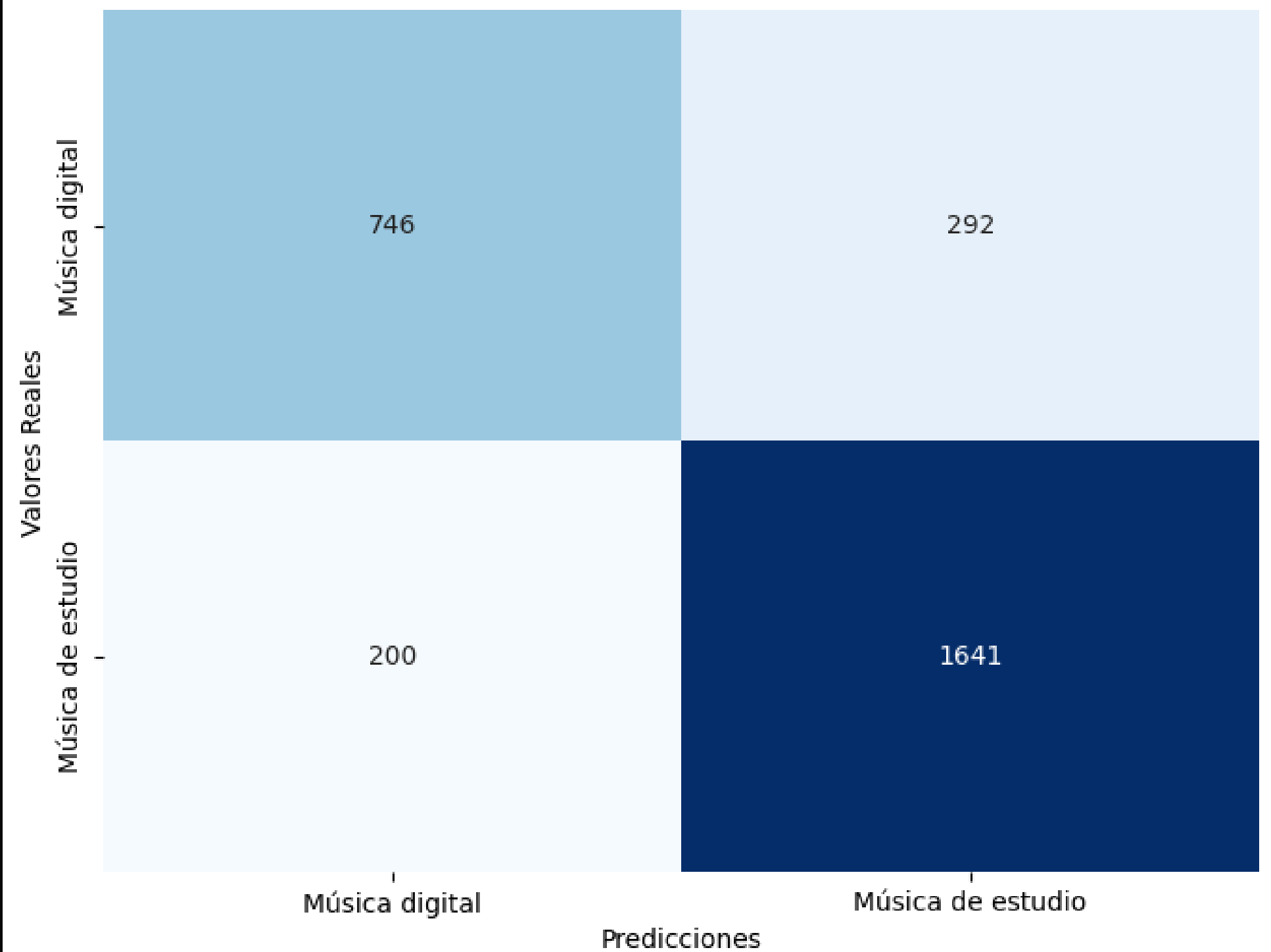


Accuracy: 0.83

Distribución de Géneros



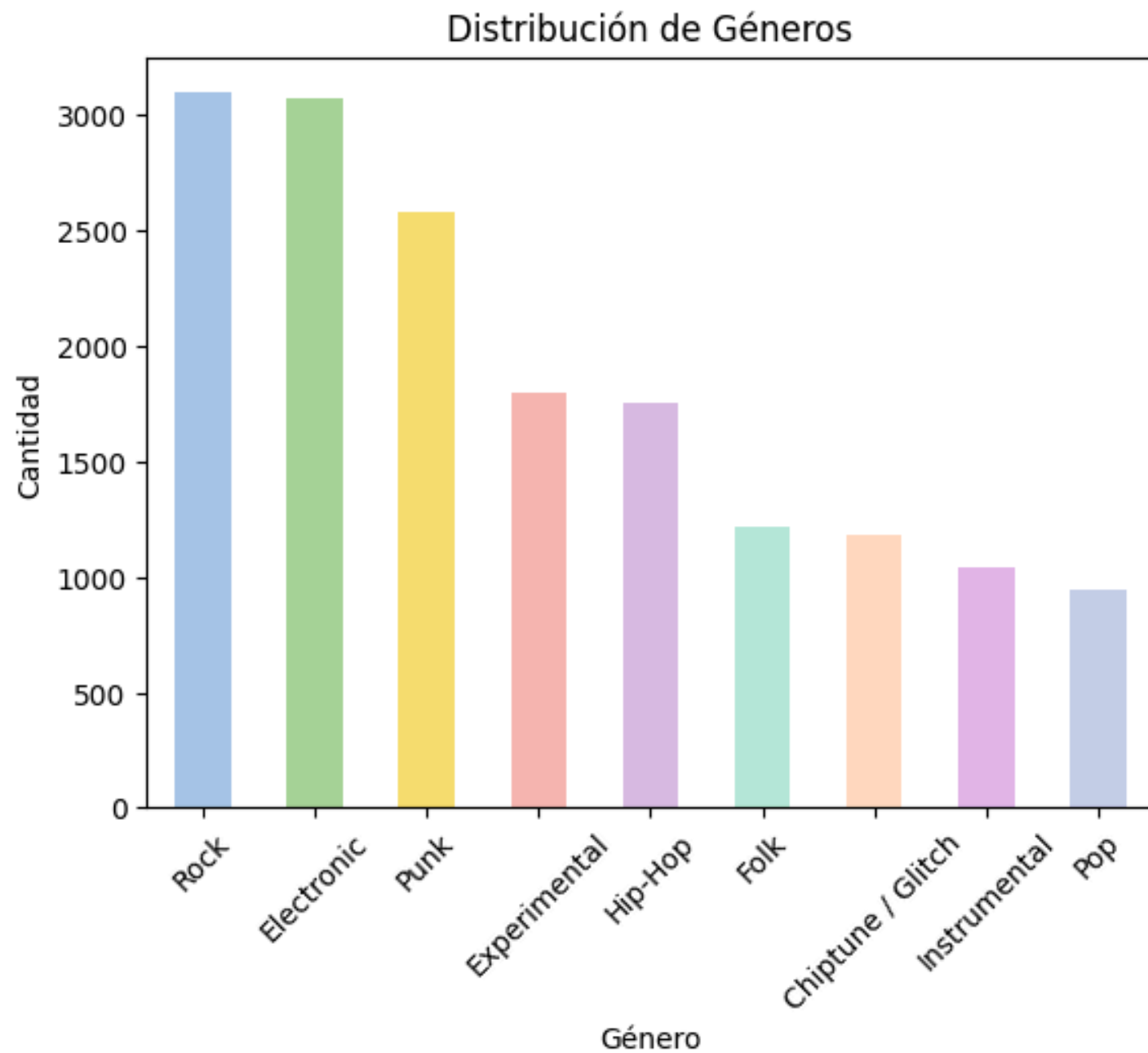
Matriz de Confusión - CatBoost



APROXIMACIÓN 1 - XGBoost



Accuracy: 0.47



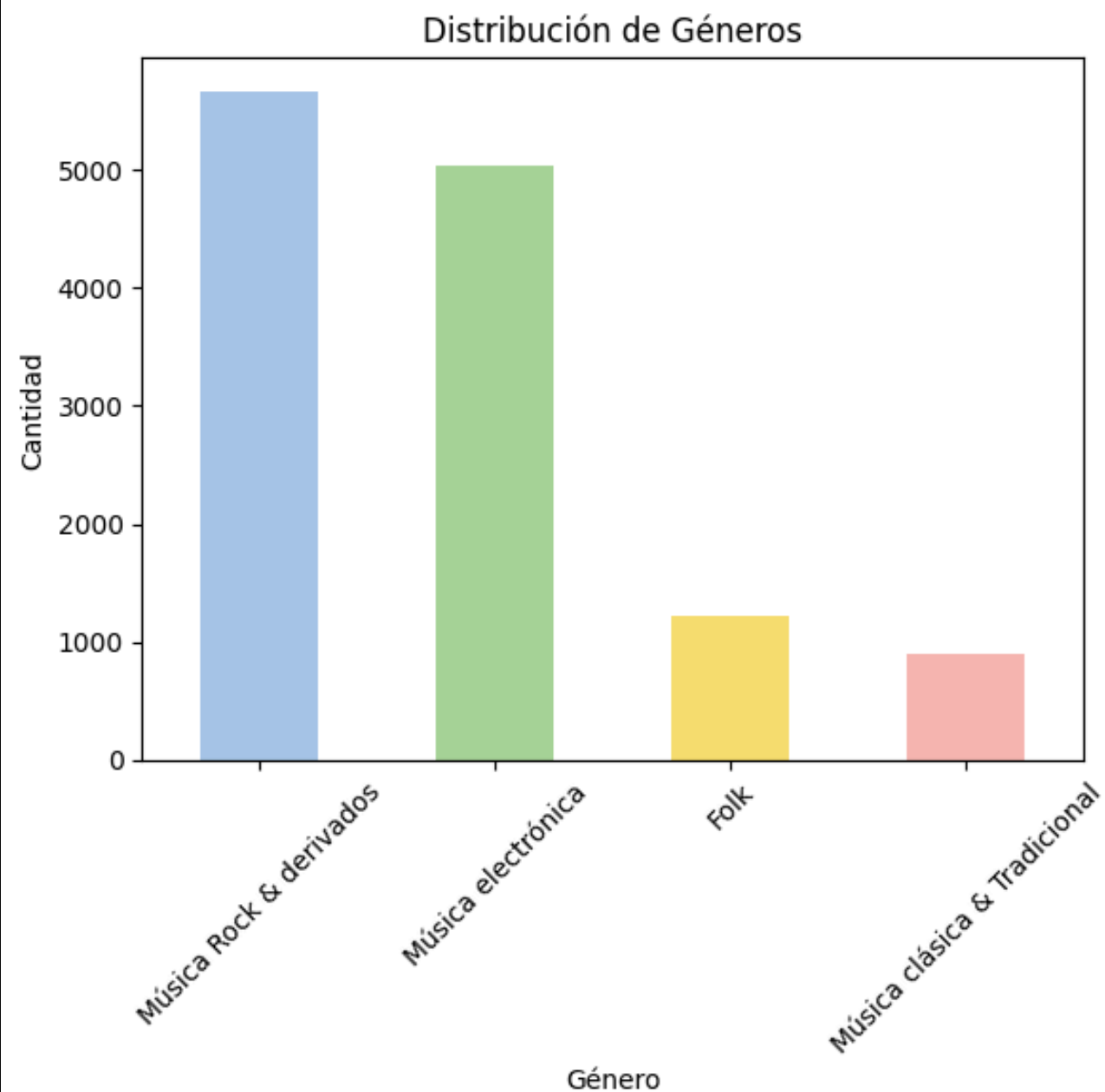
Matriz de Confusión - XGBoost

	Electronic	Rock	Punk	Experimental	Hip-Hop	Folk	Chiptune / Glitch	Instrumental	Pop
Electronic	358	50	23	43	53	17	37	11	3
Rock	64	295	109	37	24	26	6	14	4
Punk	27	141	328	17	9	9	4	4	0
Experimental	76	58	13	134	16	23	3	39	5
Hip-Hop	109	48	18	14	147	8	9	2	3
Folk	24	42	5	20	7	135	0	17	4
Chiptune / Glitch	117	23	8	13	13	2	82	1	3
Instrumental	36	32	7	40	5	12	1	69	0
Pop	36	43	19	20	15	19	4	6	18

APROXIMACIÓN 2 - LighGBM



Accuracy: 0.80



Matriz de Confusión - LightGBM

	Música Electrónica	Musica Rock & derivados	Música Clásica & Tradicional	Folk
Música Electrónica	828	134	23	26
Musica Rock & derivados	150	928	4	25
Música Clásica & Tradicional	24	8	165	3
Folk	53	60	6	131

Valores Reales

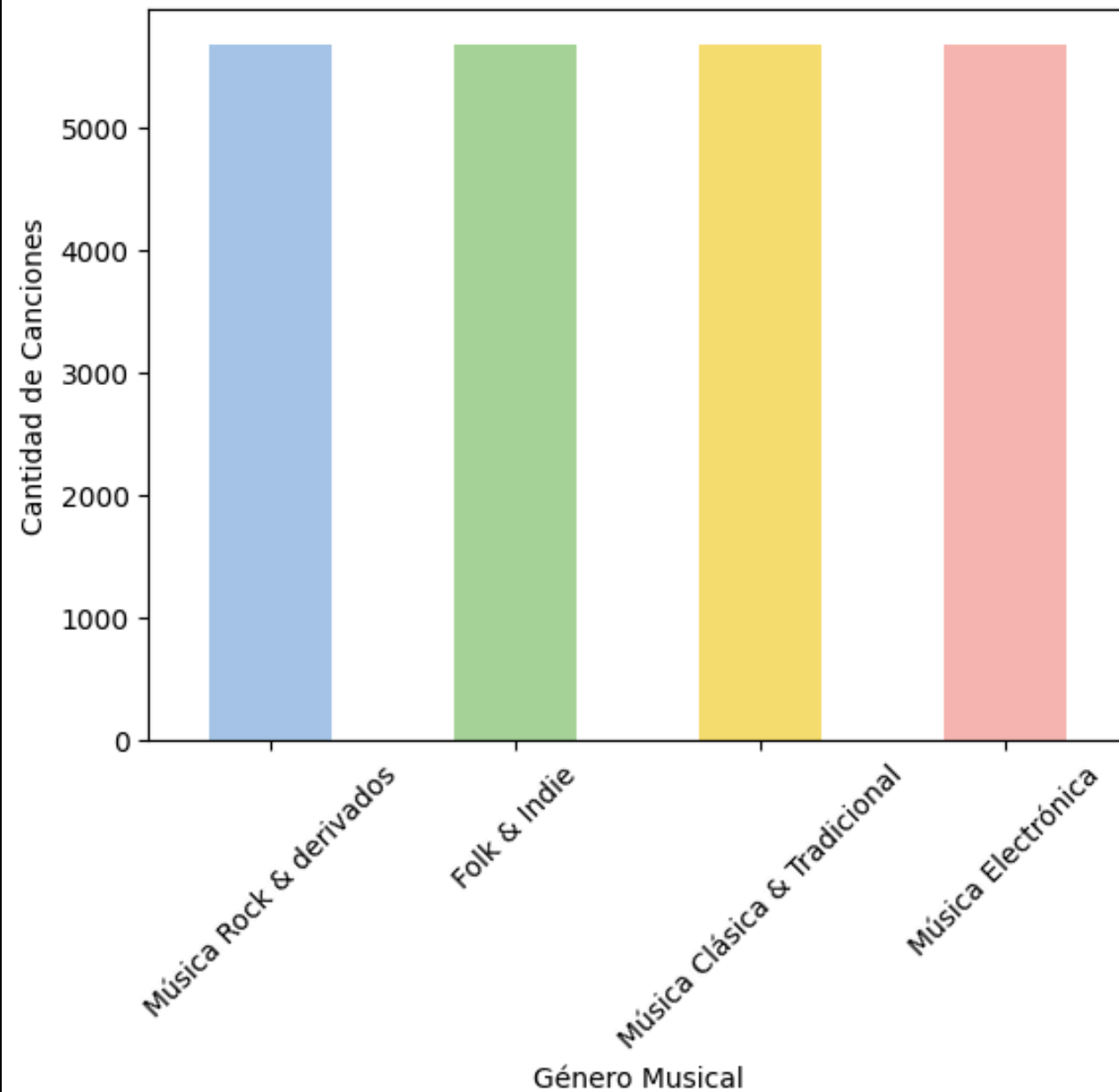
Predicciones

APROXIMACIÓN 3 - XGBoost



Accuracy: 0.88

Distribución de Géneros en el Dataset



Matriz de Confusión - XGBoost oversampled

Valores Reales					
	Música Electrónica	Musica Rock & derivados	Música Clásica & Tradicional	Folk	
	Música Electrónica	993	116	23	62
	Musica Rock & derivados	164	883	13	68
	Música Clásica & Tradicional	11	5	1096	1
	Folk	35	33	14	1023
		Predicciones			
		Música Electrónica	Musica Rock & derivados	Música Clásica & Tradicional	Folk

CONSIDERACIONES FINALES

Limitaciones

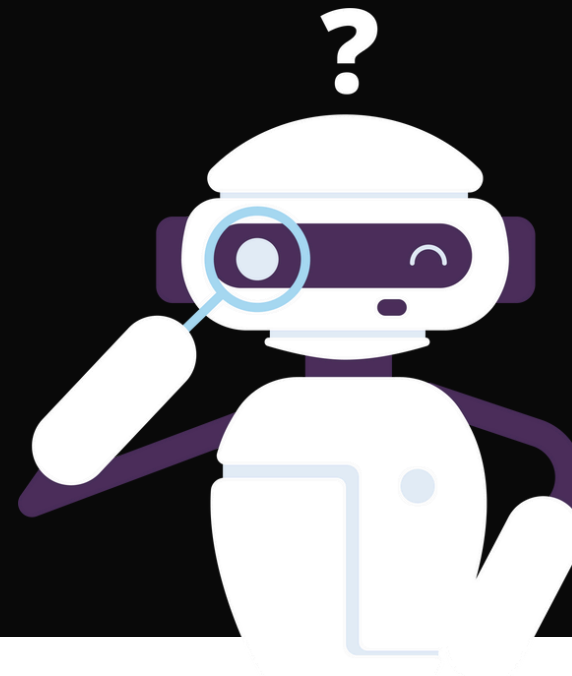
- Cantidad de datos de entrenamiento
- Complejidad de las características sonoras

Potenciales mejoras

- Aumentar el número de pistas de audio clasificadas
- Modelos más complejos
- Transfer Learning

Retos futuros

- Recolección de datos adicionales
- Costo computacional
- Diversidad cultural y evolución musical





Muchas Gracias

Construcción y Evaluación de un Modelo de Clasificación Musical



Lucas Herranz