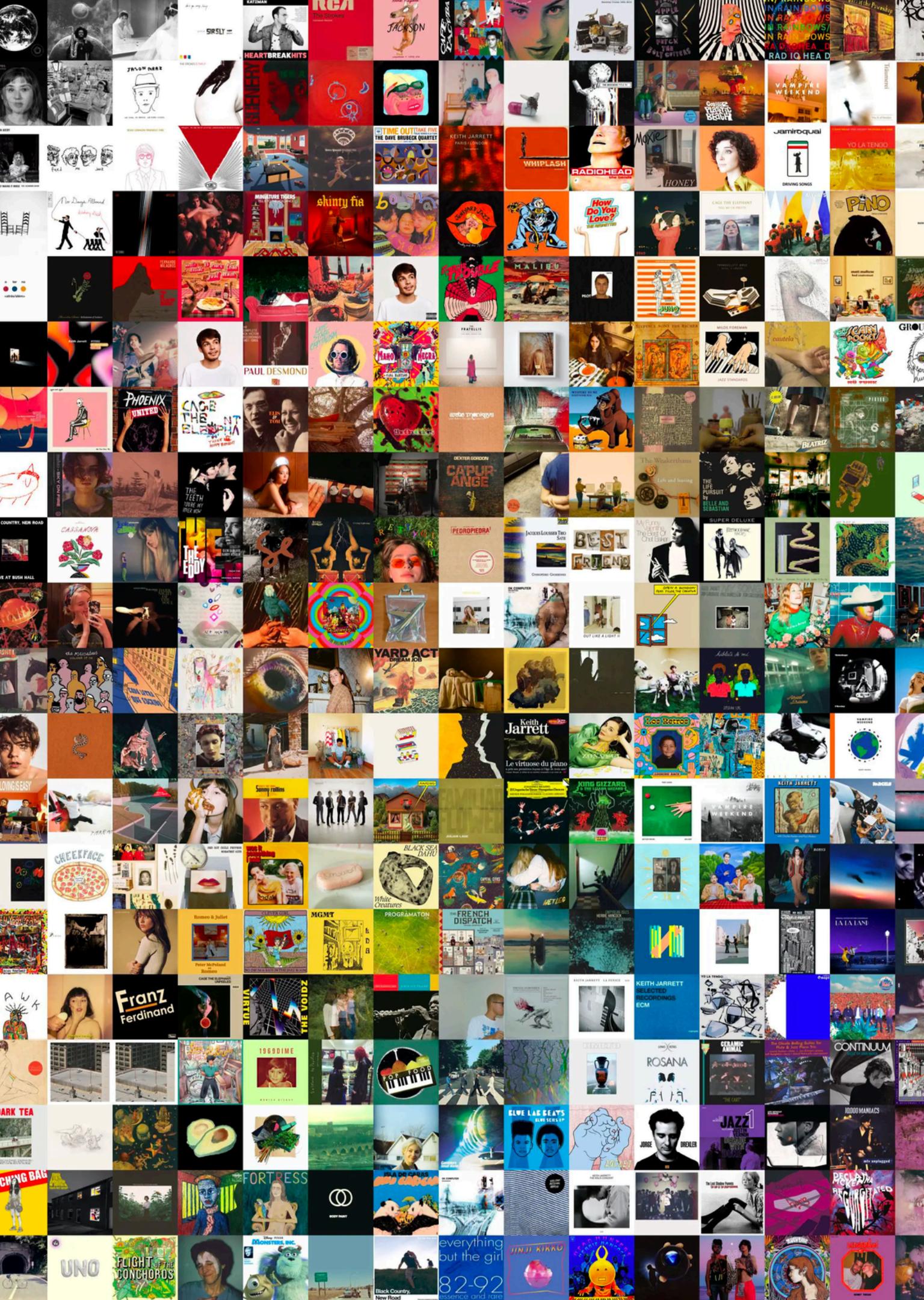


Music Taste Classifier

Santiago Villaseñor Ramírez 201427

Mauricio Díaz Villareal 200854

Jorge Esteban Ramírez Sashida 201530



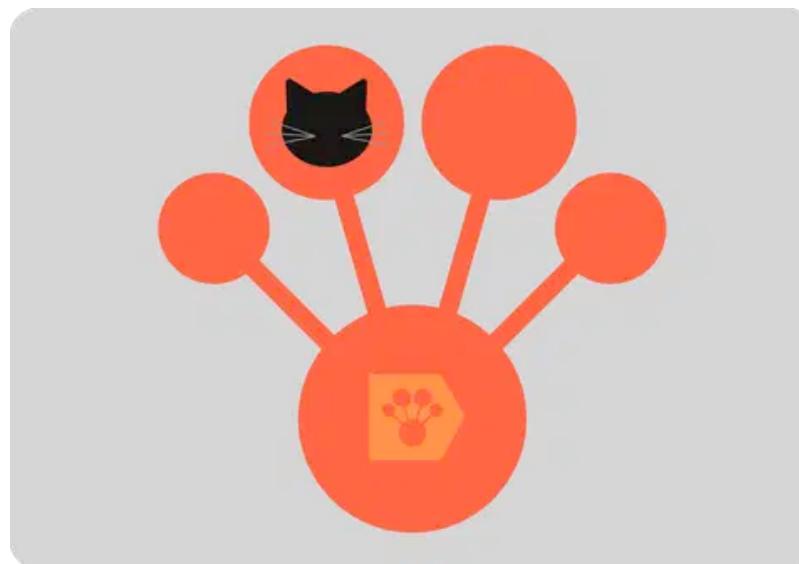
Contenido

1 Descripción del proyecto	2 Descripción de variables	3 Resumen de los datos
4 Correlaciones	5 Modelos baseline	6 Pasos siguientes

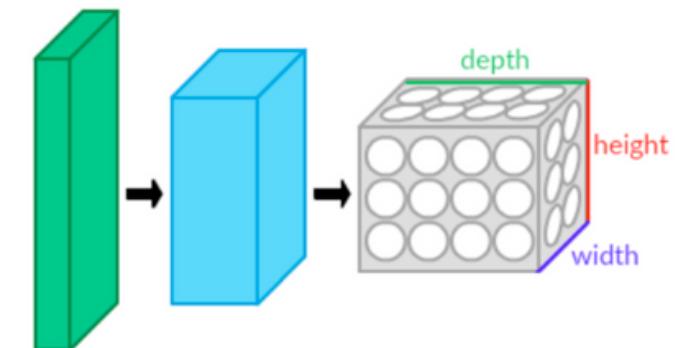
Descripción del Proyecto

Contrastar desempeño entre clasificadores usando features, audiofeatures extraido por spotify, versus clasificadores usando imágenes (espectrogramas).

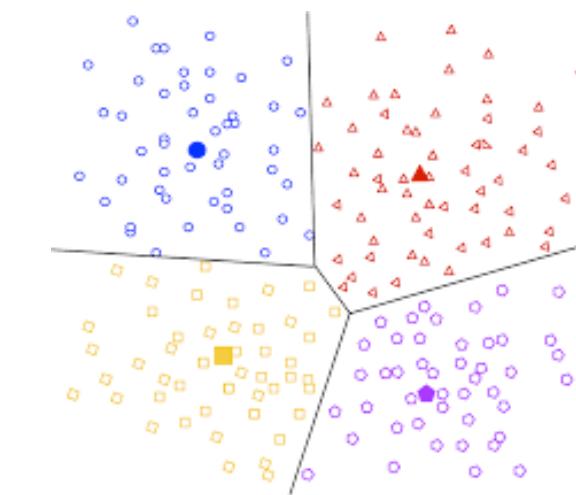
Catboost



CNN



Clustering



Descripción de variables

Tabular

Variable	Tipo	Descripción
acousticness	numérica	Qué tan acústica es la canción. (0, 1)
danceability	numérica	Qué tan apta para bailar es (tempo, estabilidad del ritmo, fuerza del beat). (0, 1)
energy	numérica	Percepción de intensidad y actividad. (0, 1)
key	categórica	Escala musical, ej:0=C, 1=C#/D♭, 2=D. (-1, 11)
liveness	numérica	Presencia de audiencia durante grabación. (0, 1)
loudness	numérica	Promedio de decibeles relativo a otras canciones. (-60, 0)

Descripción de variables

Tabular

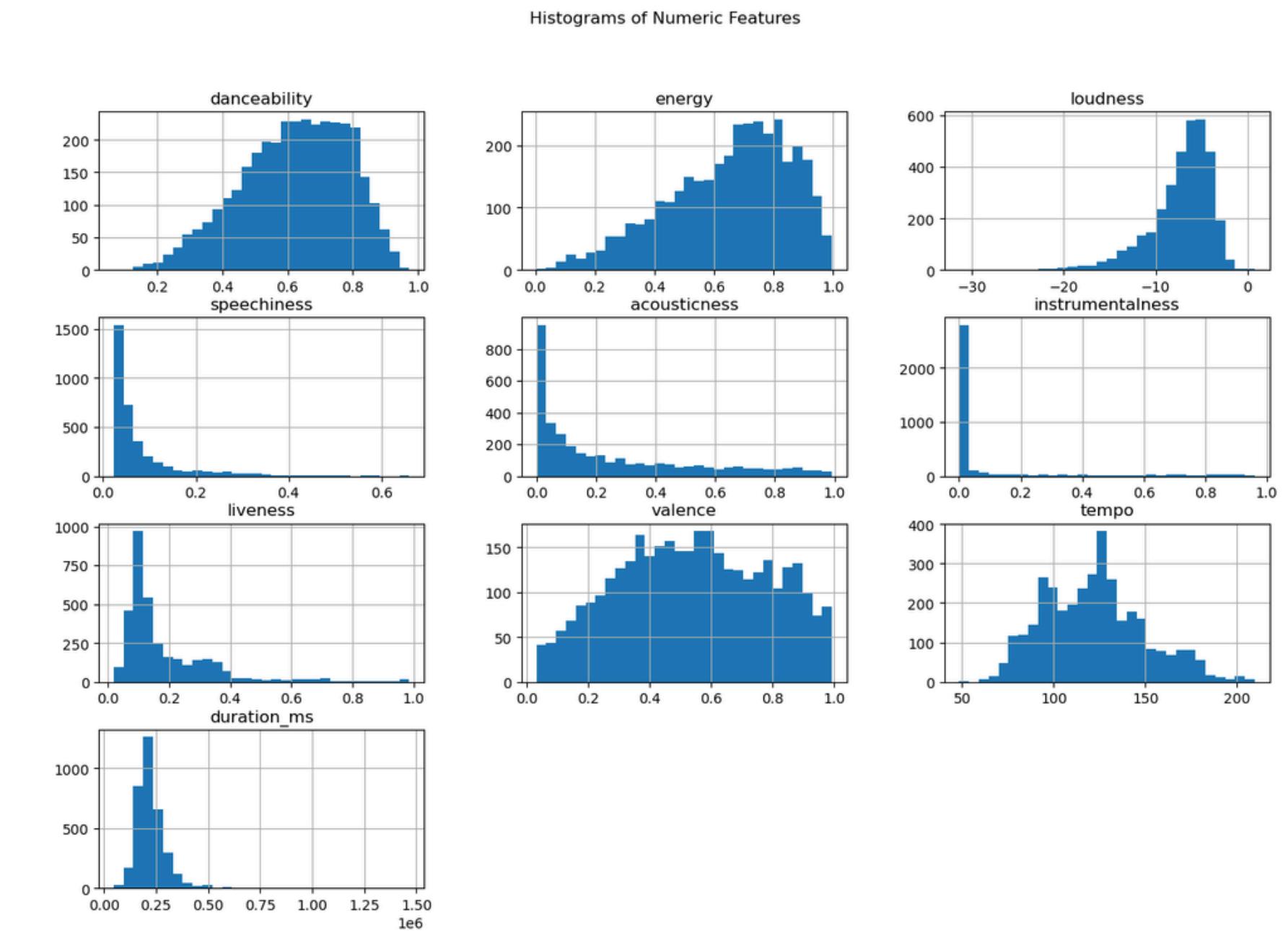
Variable	Tipo	Descripción
mode	binaria	Si está en escala menor (0) o mayor (1). {0, 1}
speechness	numérica	Presencia de palabras. (0, 1)
tempo	numérica	Beats por minuto. (50, 200)
time_signature	ordinal	Tipo de compás (todas sobre 4). {3, 4, 5, 6, 7}
valence	numérica	Qué tan “feliz” o “positiva” es la canción. (0, 1)

Resumen de los datos

	count	mean	std	min	25%	50%	75%	max
acousticness	3420	0.255841	0.276937	0.000002	0.028375	0.139000	0.425000	0.992
danceability	3420	0.619656	0.162903	0.066500	0.506750	0.633000	0.748000	0.973
duration_ms	3420	226542.365205	82780.628828	46227.000000	181860.250000	214793.500000	253003.500000	1465973.000
energy	3420	0.651769	0.204109	0.003790	0.515000	0.687000	0.811000	0.997
instrumentalness	3420	0.084640	0.214206	0.000000	0.000000	0.000062	0.011200	0.961
key	3420	5.370468	3.550645	0.000000	2.000000	5.000000	9.000000	11.000
like	3420	0.505263	0.500045	0.000000	0.000000	1.000000	1.000000	1.000
liveness	3420	0.190874	0.160283	0.019100	0.096075	0.126000	0.239250	0.986
loudness	3420	-7.210553	3.370302	-31.414000	-8.715250	-6.516000	-4.871000	0.811
mode	3420	0.632749	0.482126	0.000000	0.000000	1.000000	1.000000	1.000
speechiness	3420	0.078560	0.076332	0.022900	0.034800	0.047900	0.085525	0.660
tempo	3420	121.716892	27.628945	48.349000	99.974750	120.915000	138.035000	209.688
time_signature	3420	3.921345	0.330670	1.000000	4.000000	4.000000	4.000000	5.000
valence	3420	0.540879	0.240567	0.032000	0.357000	0.542500	0.739000	0.993

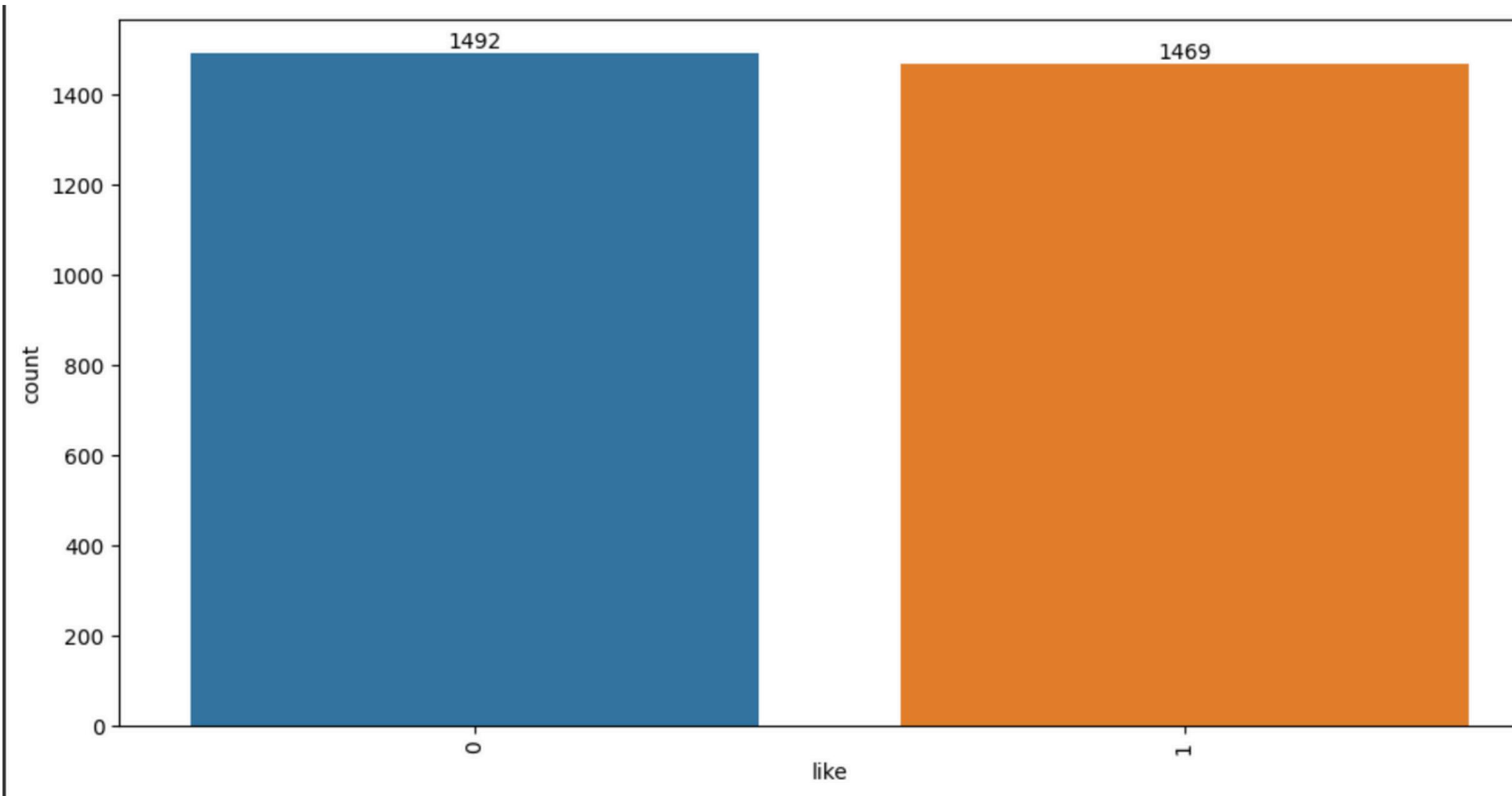
Resumen de los datos

Outlier analysis:		
	Outlier Count	Percentage
danceability	3	0.09
energy	4	0.11
key	0	0.00
loudness	52	1.49
mode	0	0.00
speechiness	92	2.64
acousticness	0	0.00
instrumentalness	160	4.59
liveness	101	2.90
valence	0	0.00
tempo	6	0.17
duration_ms	45	1.29
time_signature	33	0.95
like	0	0.00



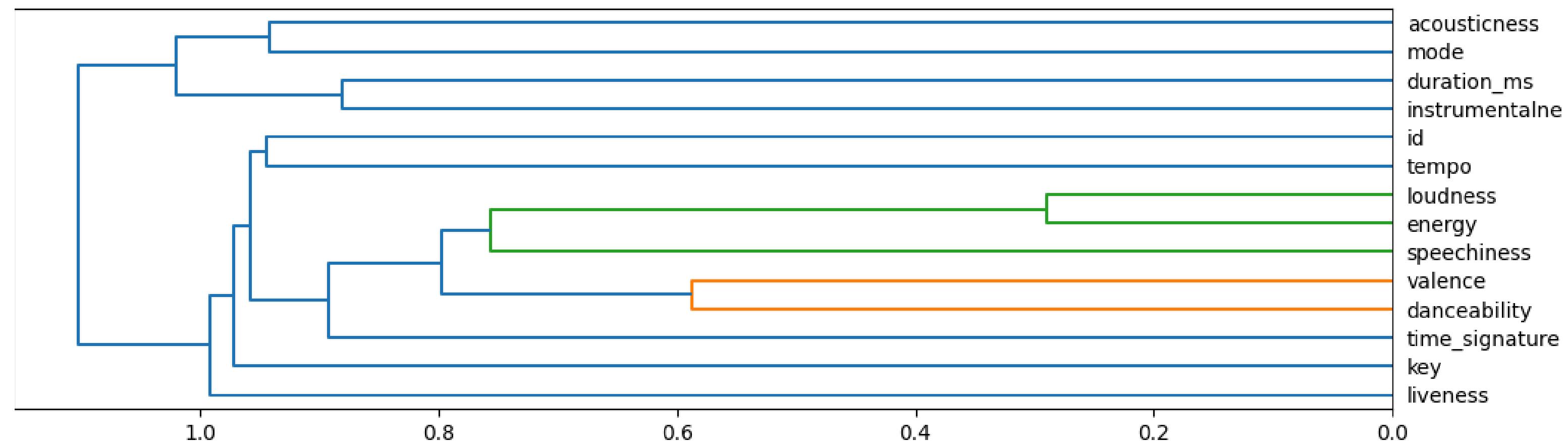
Resumen de los datos

Variable Objetivo

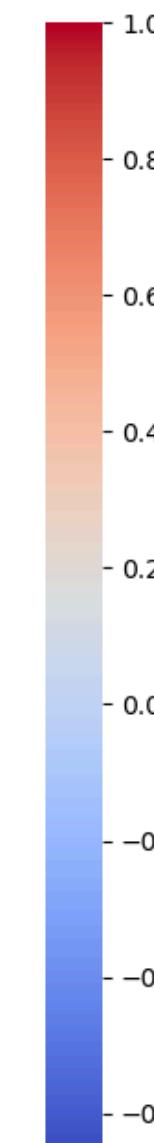
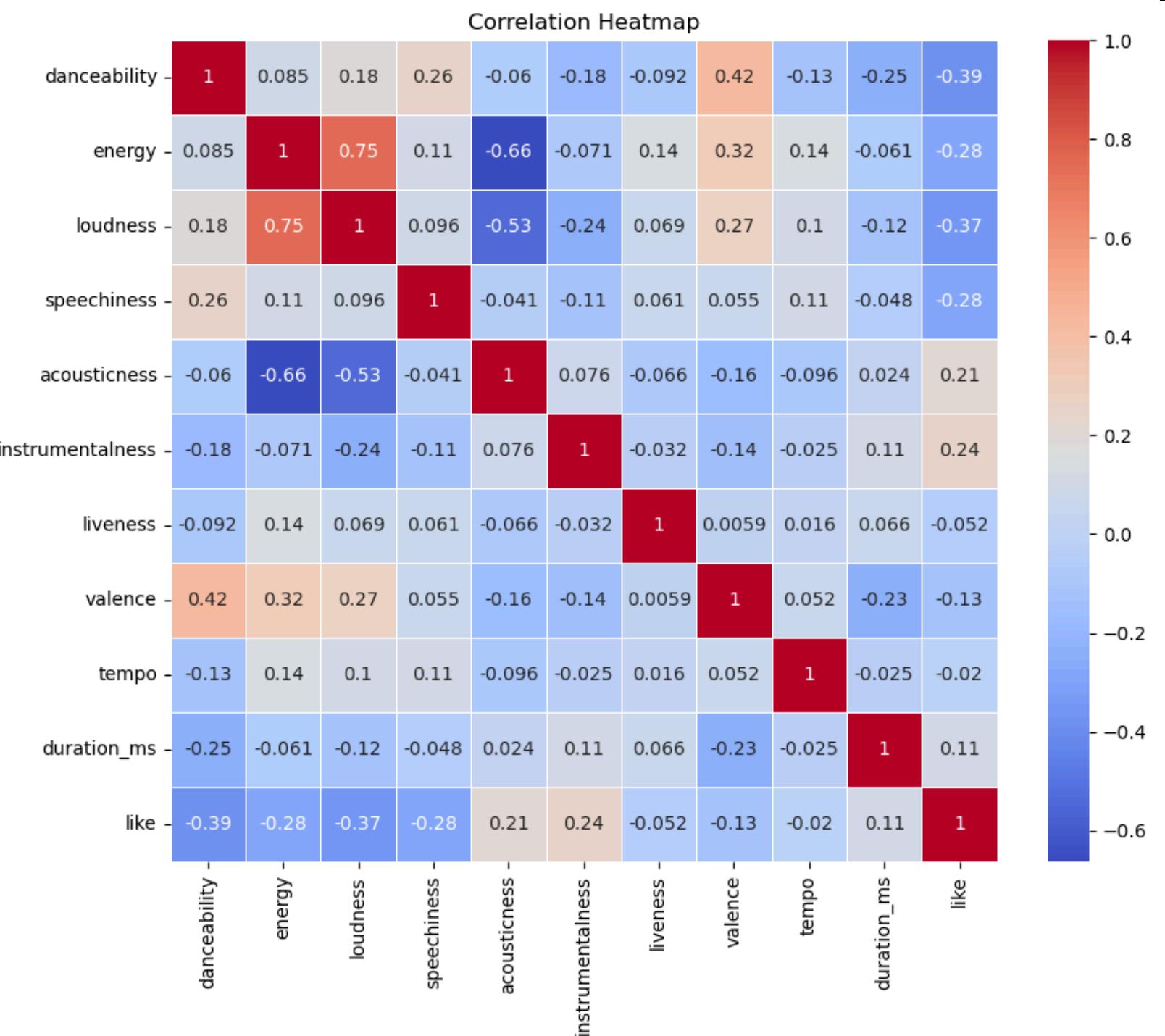


- Distribución
 - 0: 50.37%
 - 1: 49.63%

Distancia entre variables



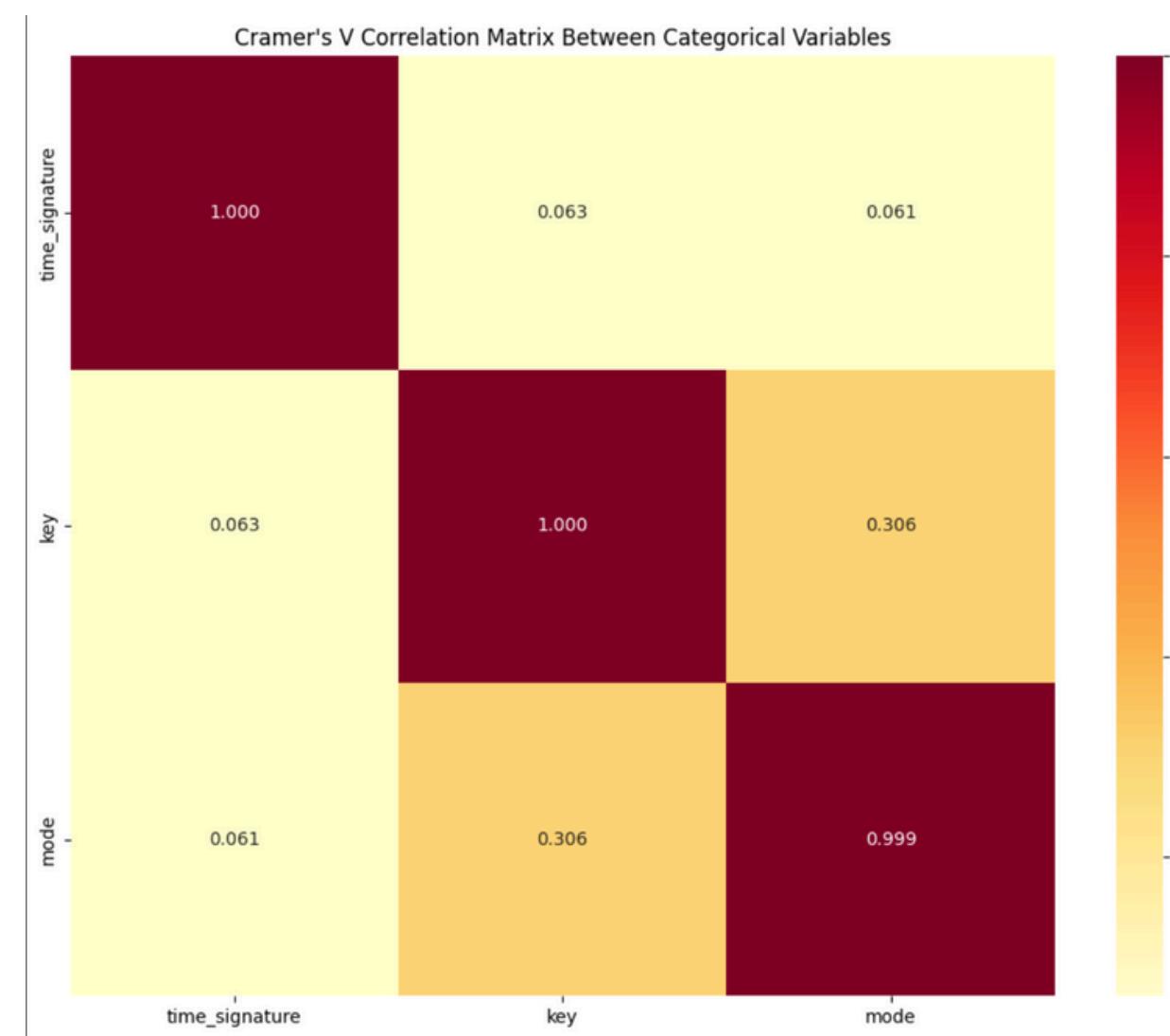
Correlaciones



- Variables altamente correlacionadas
 - Energy y acousticness
 - Loudness y energy
 - Acousticness y loudness

Correlaciones

Correlación entre categóricas

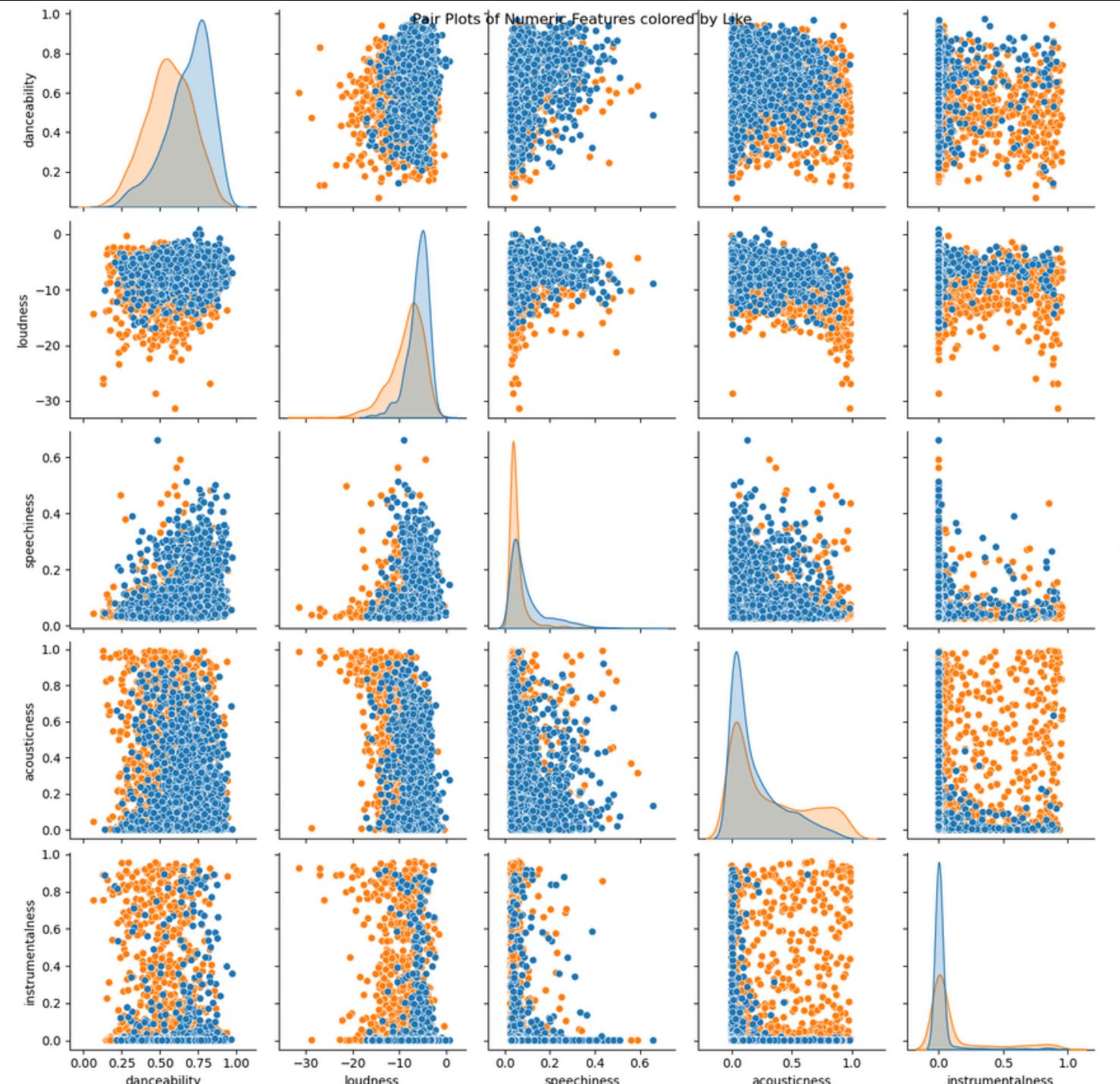


Correlación de categóricas con variable objetivo

Summary of Correlations with Target Variable:

Variable	Cramers_V	Chi_Square	P_Value
key	0.149377	77.739787	4.030003e-12
mode	0.108386	40.928294	1.579182e-10
time_signature	0.038286	5.106964	1.641302e-01

Scatterplots con target variable

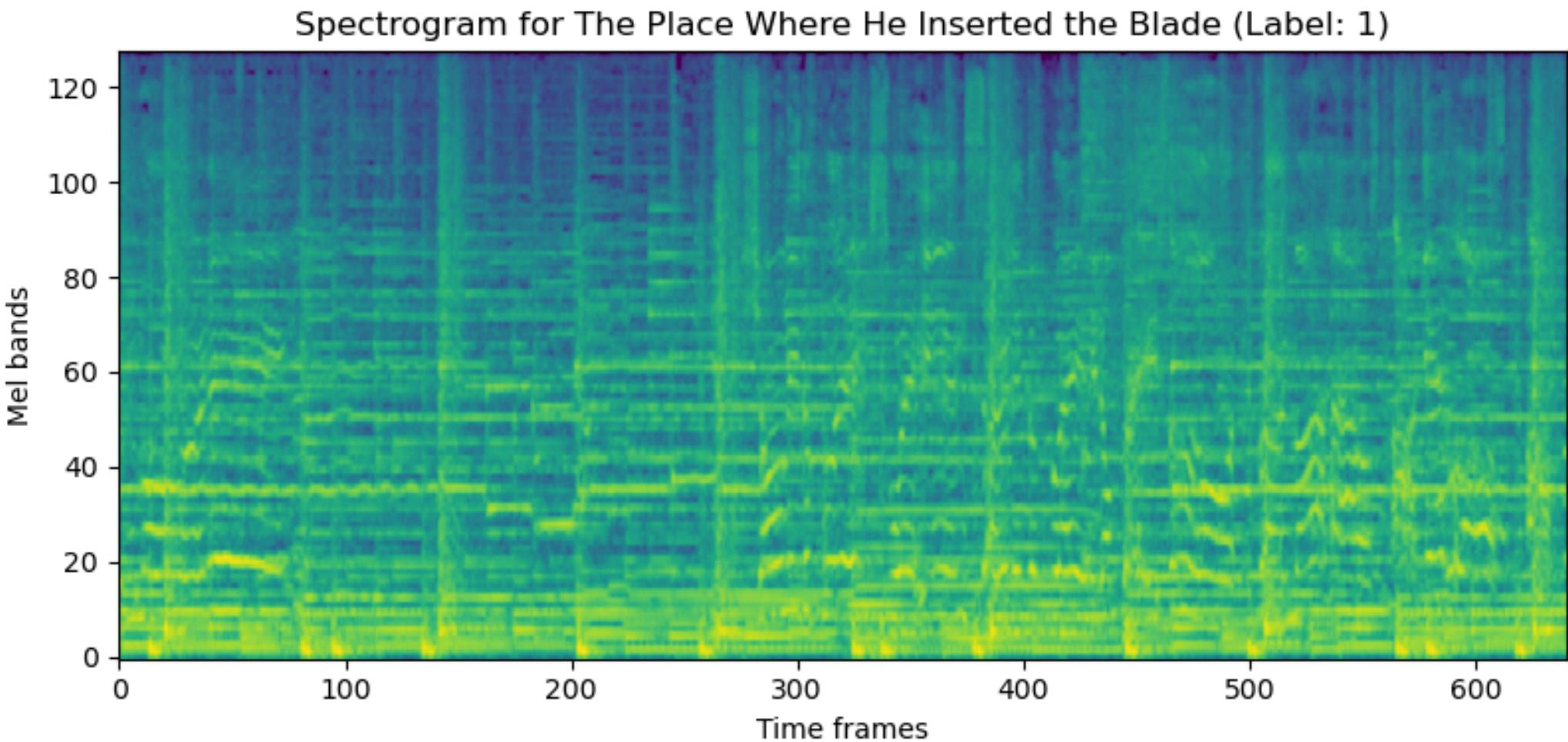
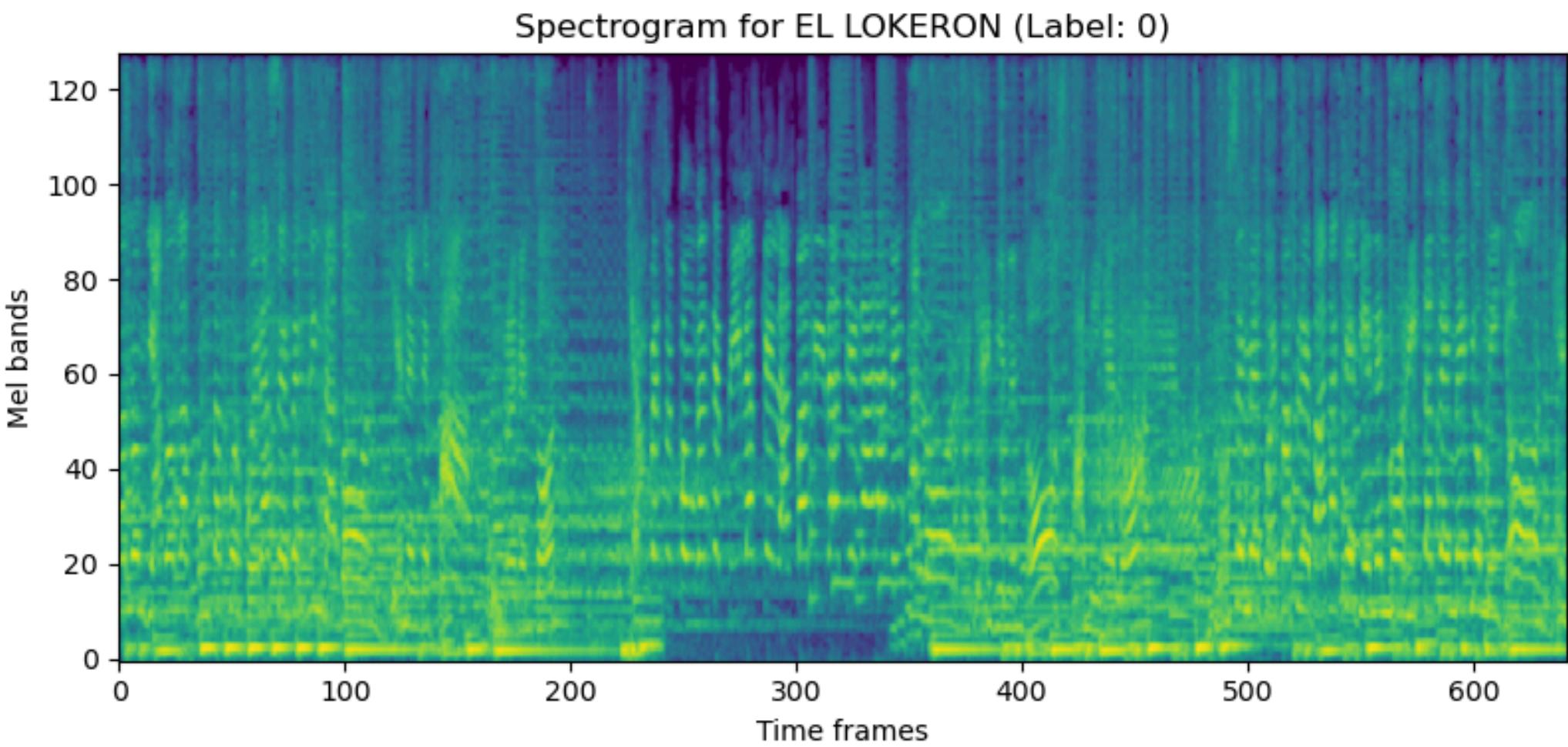


Observaciones:

- Ningún par de variables separa por completo las clases.
- Algunos pares de variables separan parcialmente
 - Instrumentalness vs acousticness.
- ¿Combinaciones de más de dos variables pueden separar?

Mel Spectrograms

- Espectrograma
 - Eje x: tiempo
 - Eje y: frecuencia
 - Intensidad de color: amplitud o energía de cada frecuencia en cada instante.
- Mel-spectrograma:
 - Las frecuencias por la escala mel.

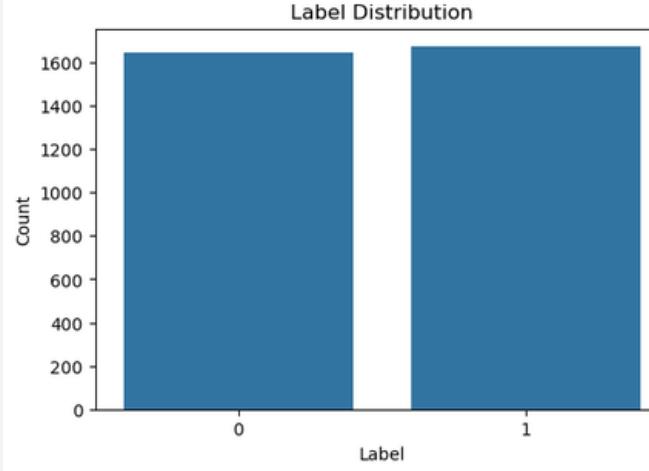


Descripción general

- Tensores (matrices) de 128×641 .
- Representan muestras de 14.90 segundos.
- Todos son del mismo tamaño

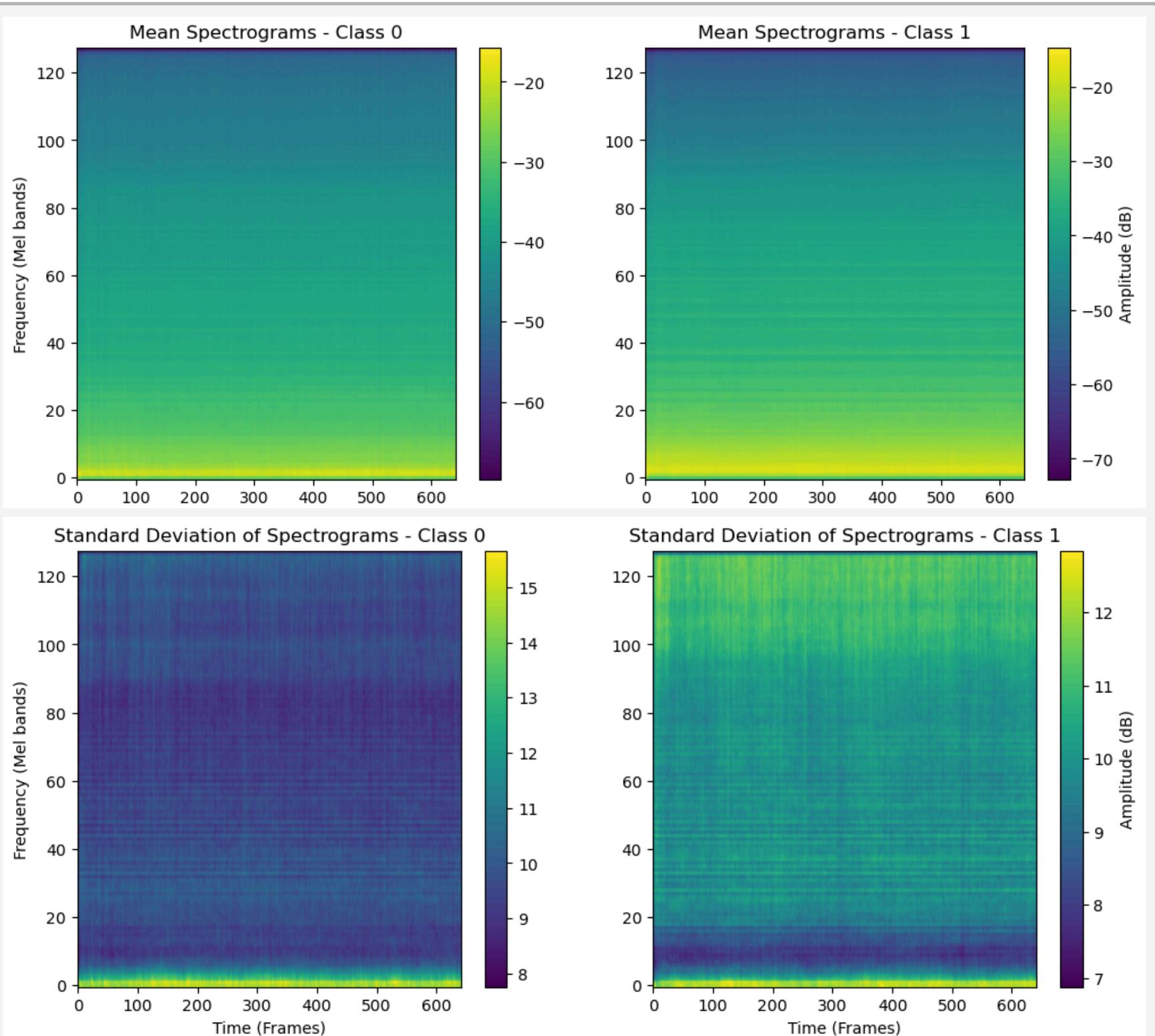
Audio a imagen

- Tamaño de salto: 512
- Tasa de sampling: 22050 Hz
- Cuadros de tiempo: 641



Datos balanceados

SPECTS EDA

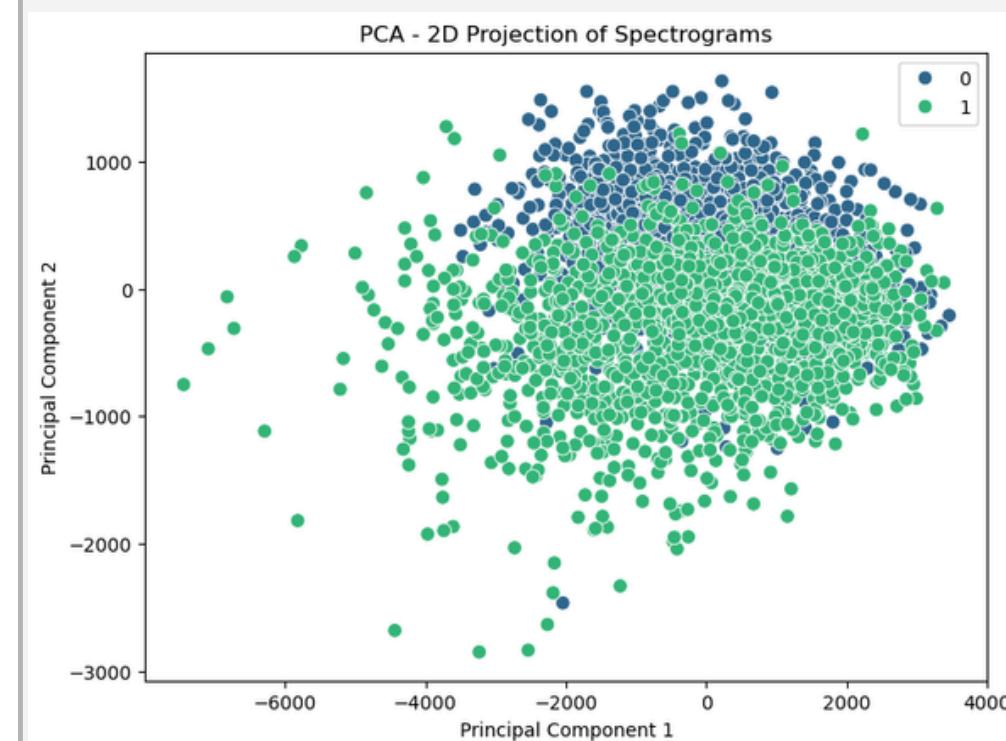


Medias

Se parecen mucho, pero hay más amplitud promedio en las bandas bajas en clase 1.

Desviaciones estándar

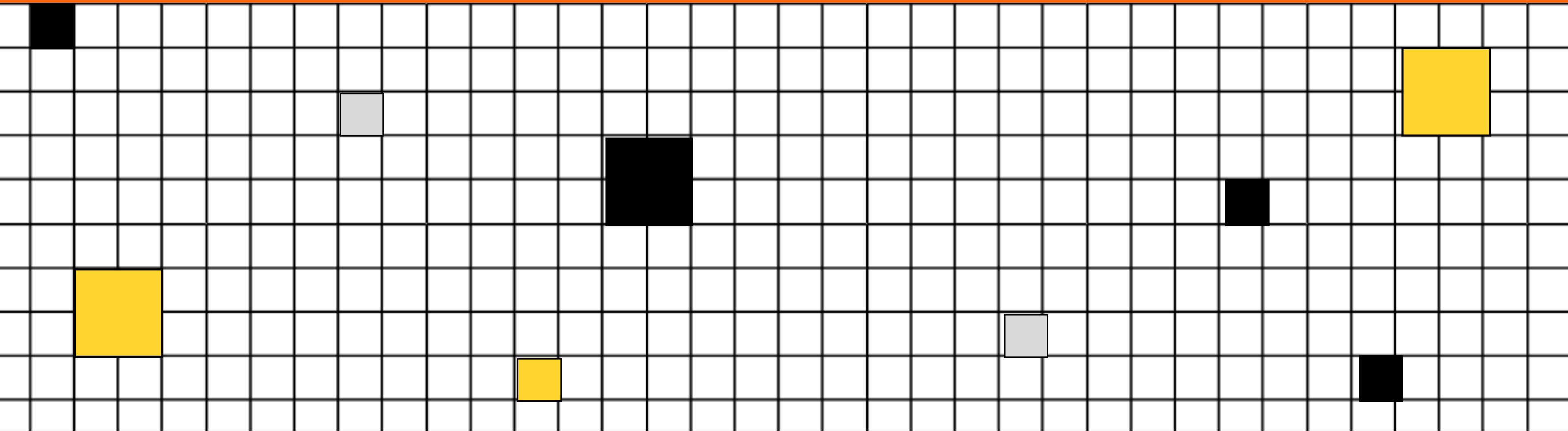
- La clase 1 varía más que la clase 0, especialmente en las bandas altas.
- La clase 0 varía más solo en las bandas más bajas.



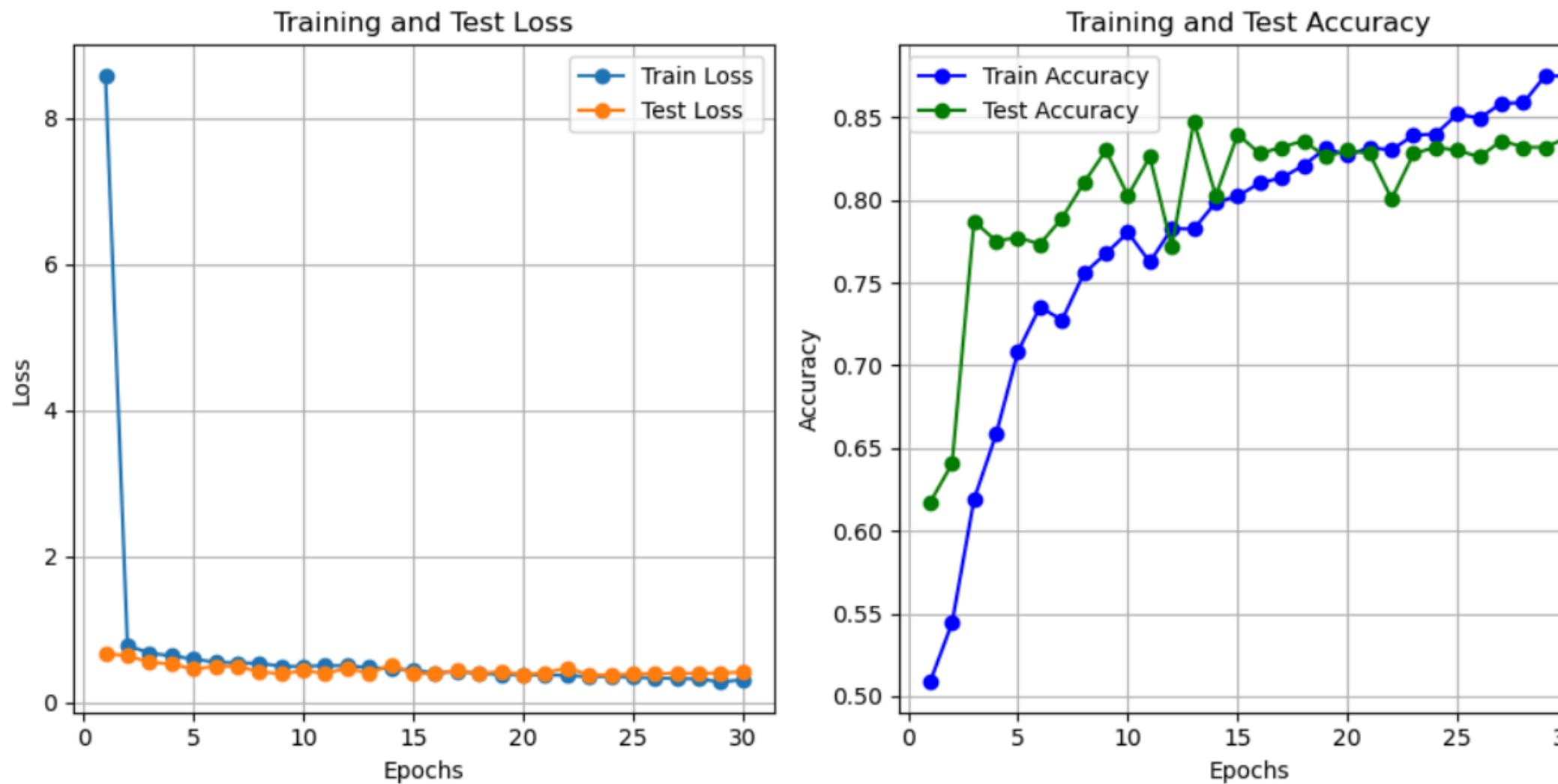


Primeros modelos

CNN y Catboost



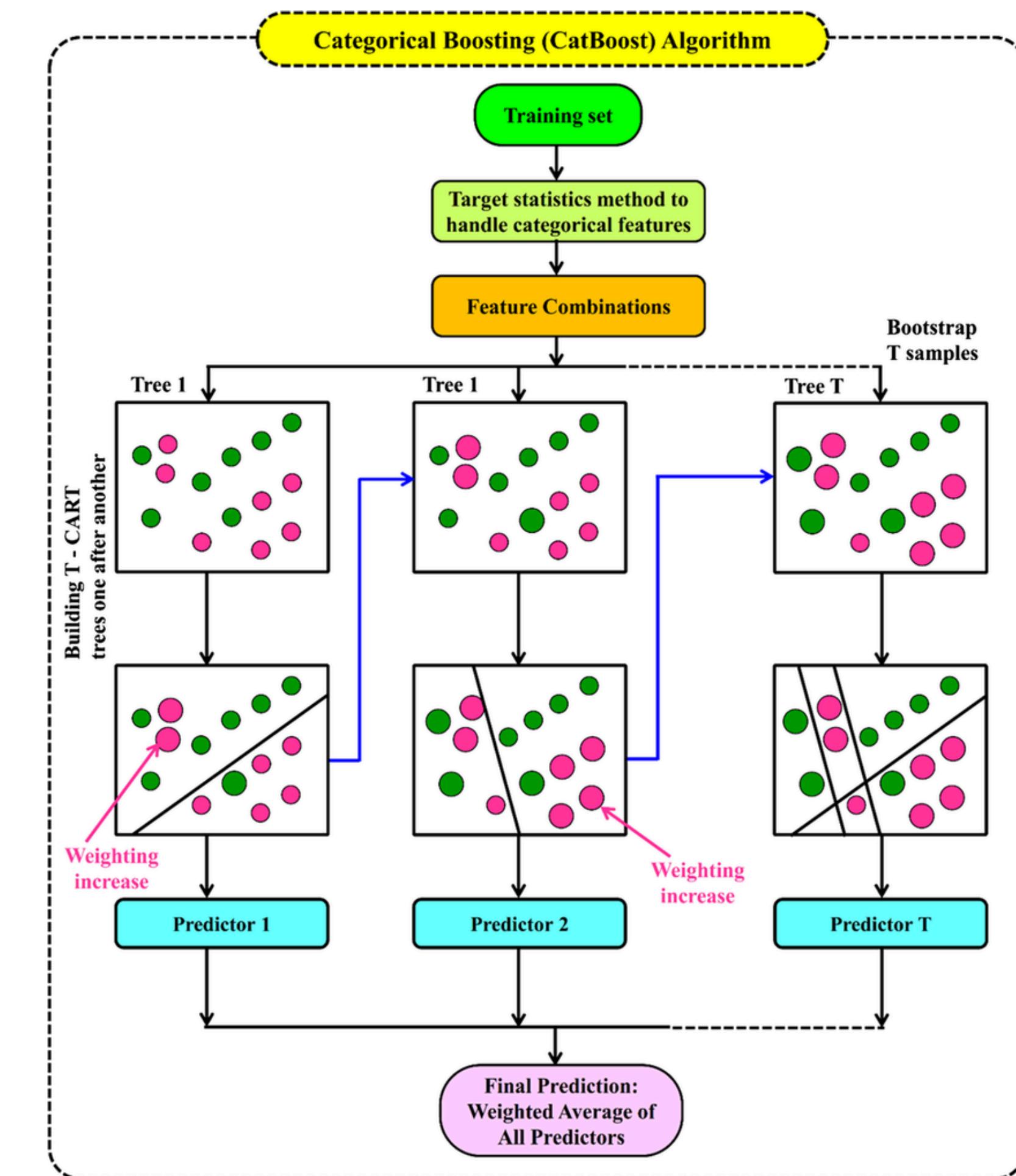
Convolutional Neural Network



- Arquitectura
 - Entrada: Imagen de tamaño $128 \times 641 \times 1$.
 - 3 Bloques Convolucionales
 - Dos capas densas con Dropout y ReLU.
 - Capa de salida: capa lineal (podríamos usar sigmoide).
- Train:
 - Loss: 0.3083
 - Accuracy: 0.8747
- Test
 - Loss: 0.4129
 - Accuracy: 0.8379

Catboost

- Modelo:
 - Modelo de boosting en árboles de decisión.
- Train:
 - Accuracy: 0.905
- Test
 - Accuracy: 0.83
 - Balanced accuracy: 0.83
 - MCC: 0.66
 - ROC: 0.9
 - Precision: 0.81
 - Recall: 0.86



¡Gracias!

