

Entrega TP3

Nombre: Alejo Valentín Fábregas

Padrón: 106160

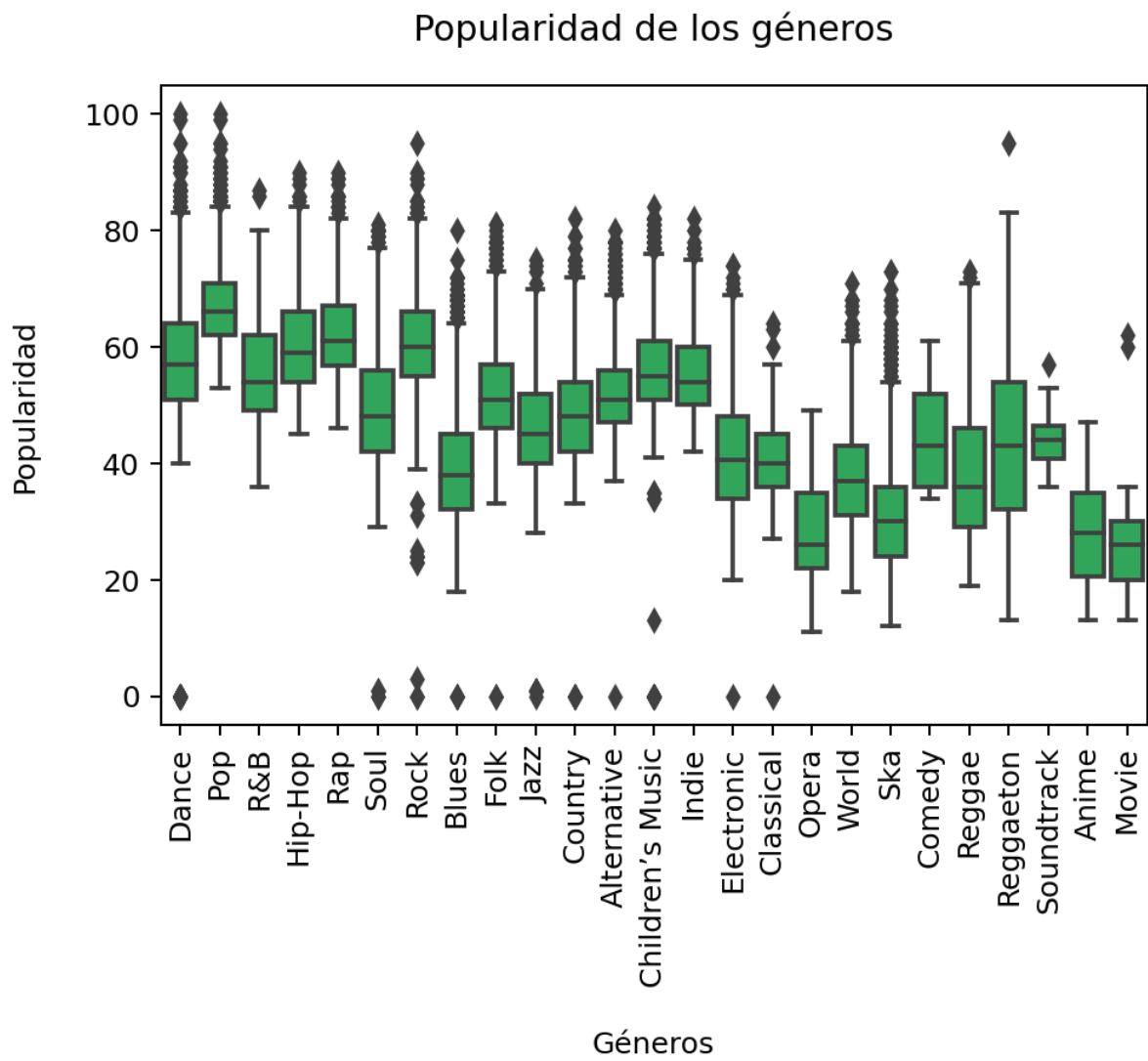
Corrector: Ignacio Brusati y Juan Pablo Fresia

Parte 1

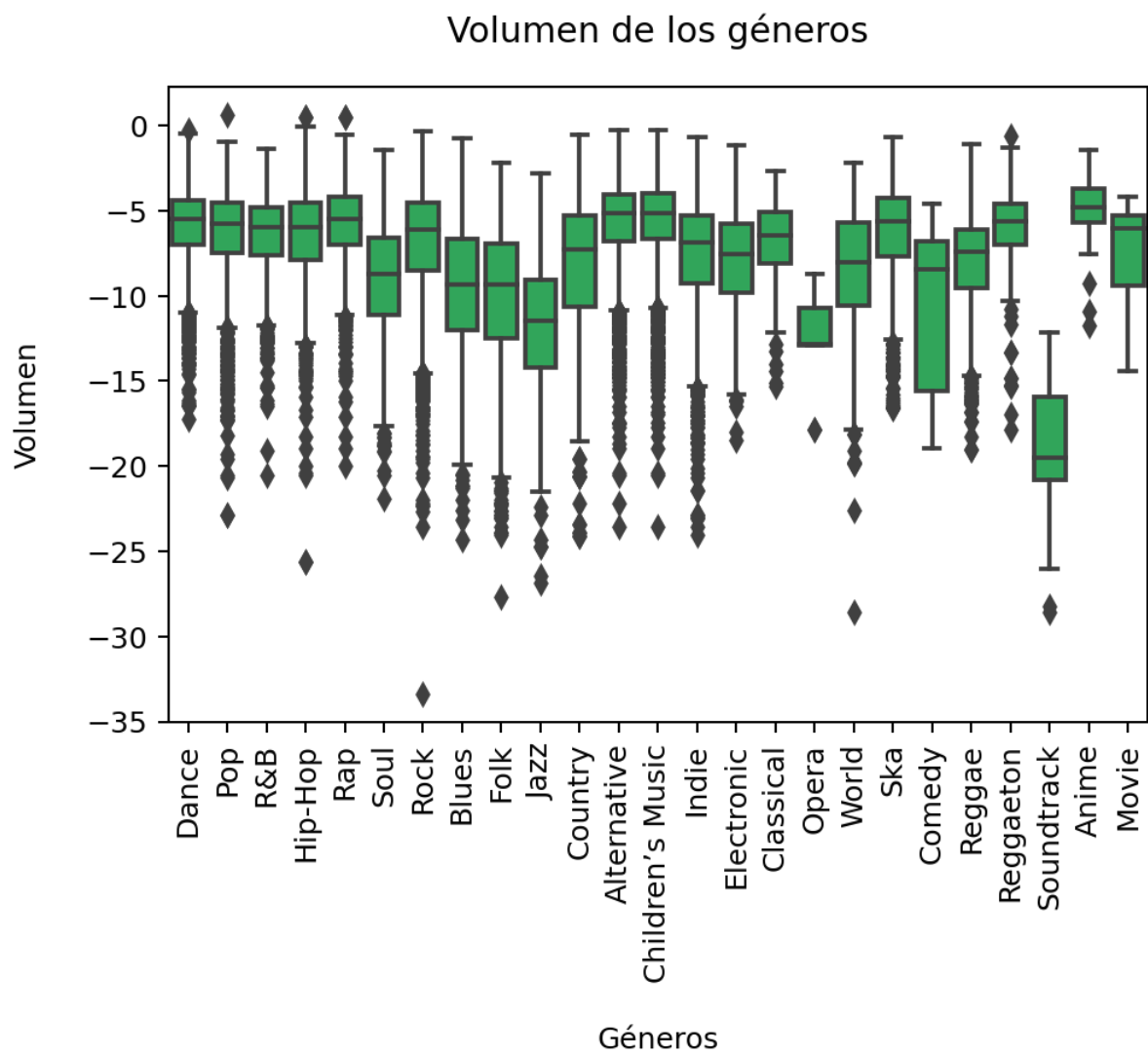
Link al notebook con todas las visualizaciones:

<https://drive.google.com/file/d/16EGHDm4lsVm9rc3DZpjDLnme4pCFiRXA/view?usp=sharing>

Visu 1

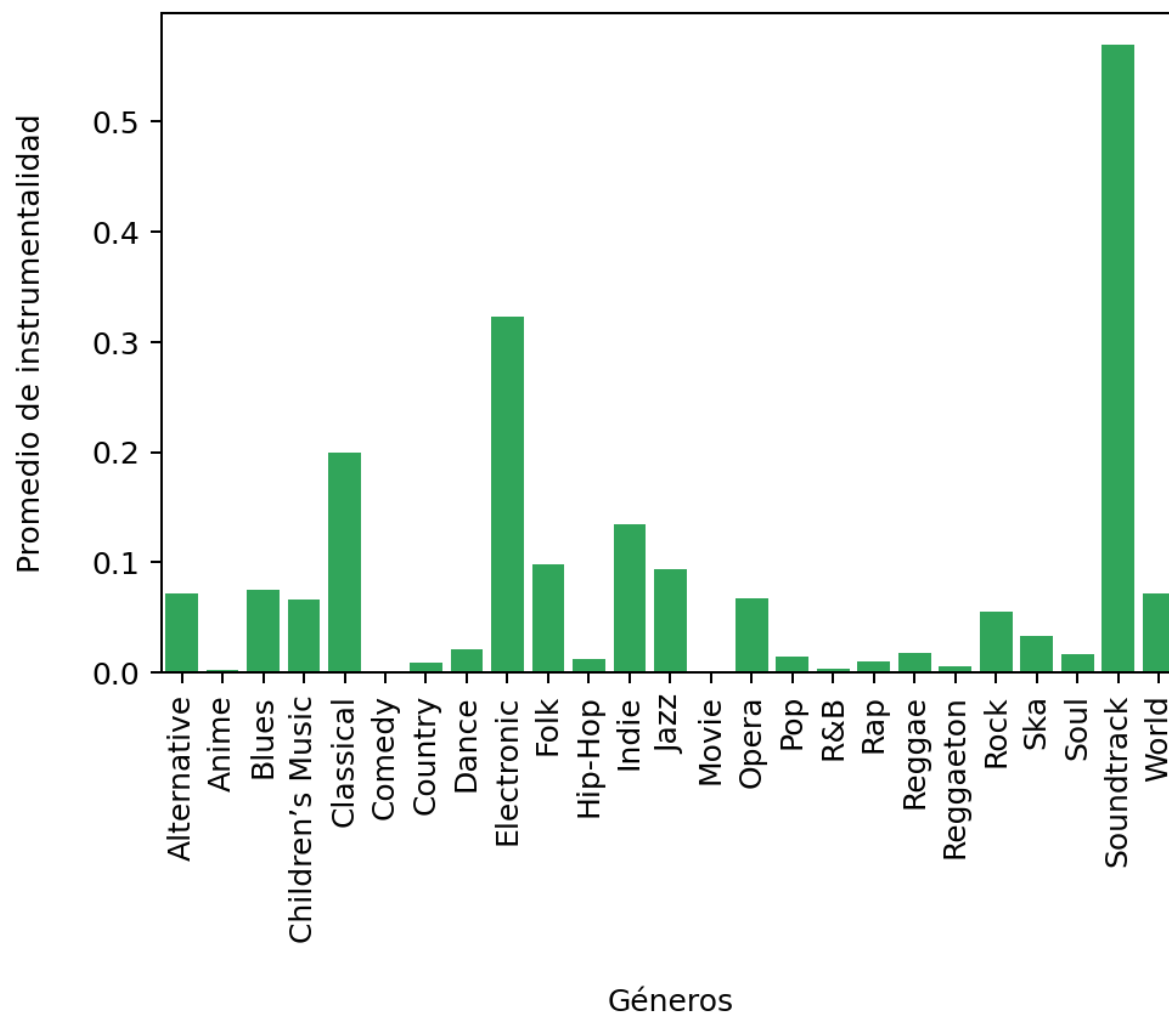


Visu 2

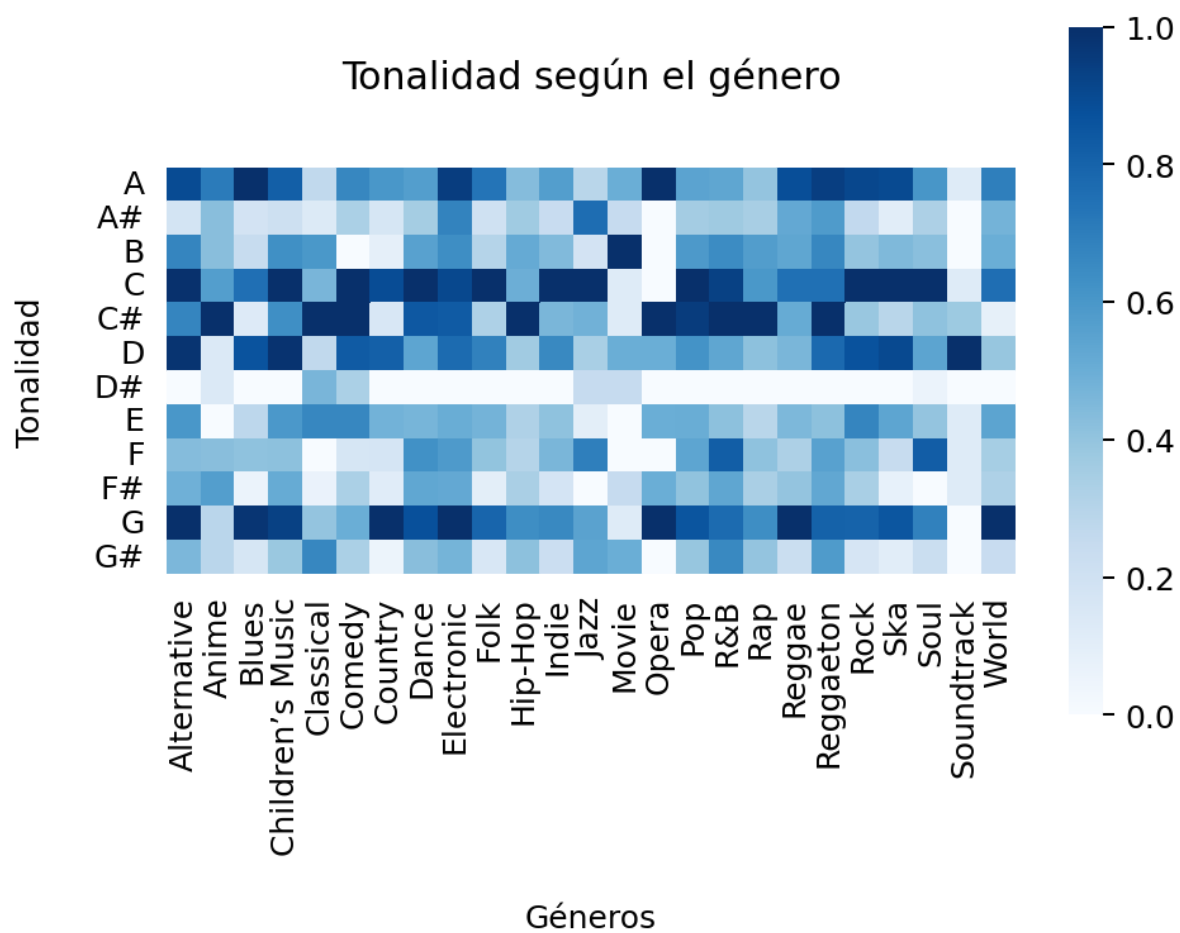


Visu 3

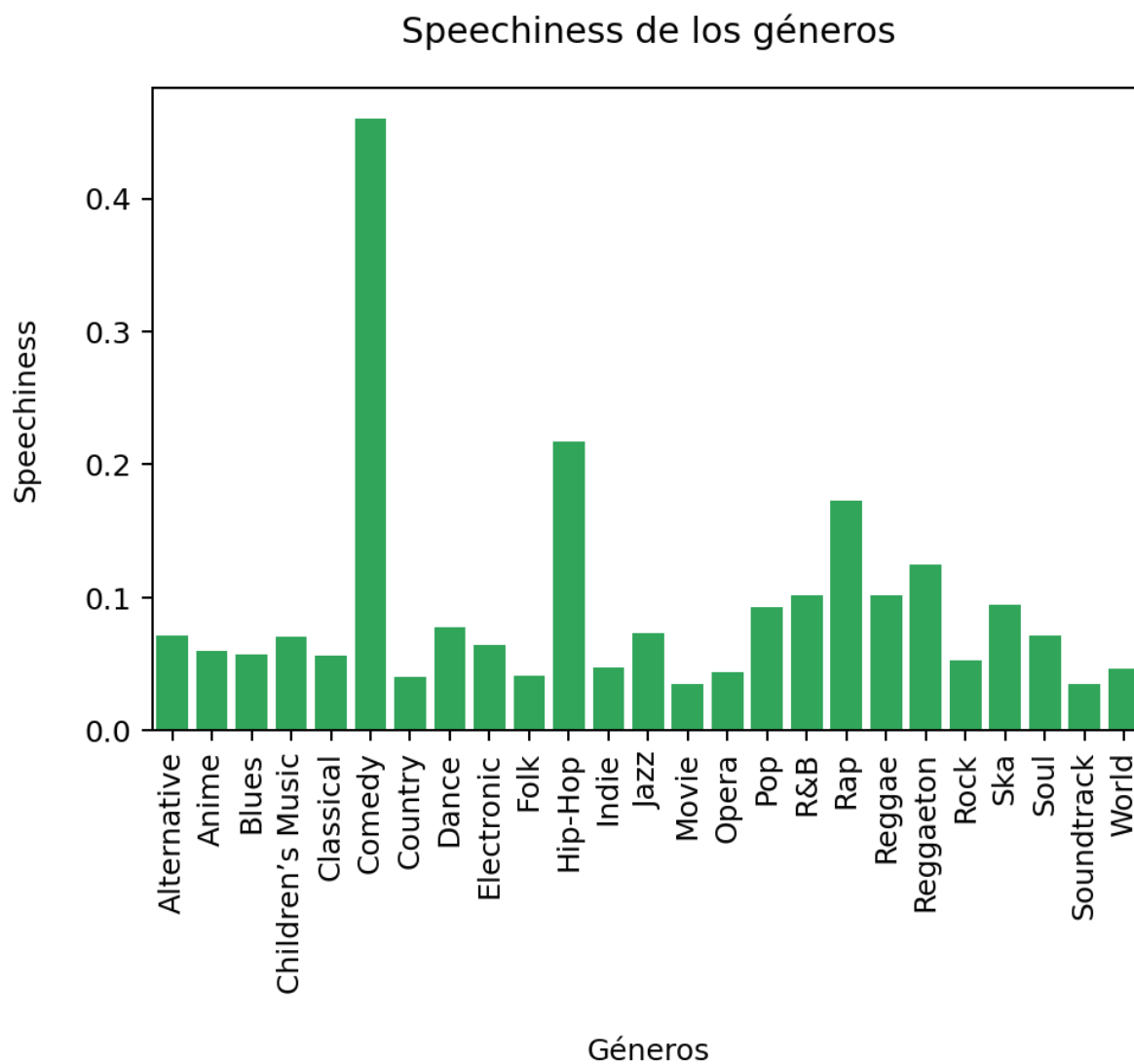
Instrumentalidad de los géneros



Visu 4

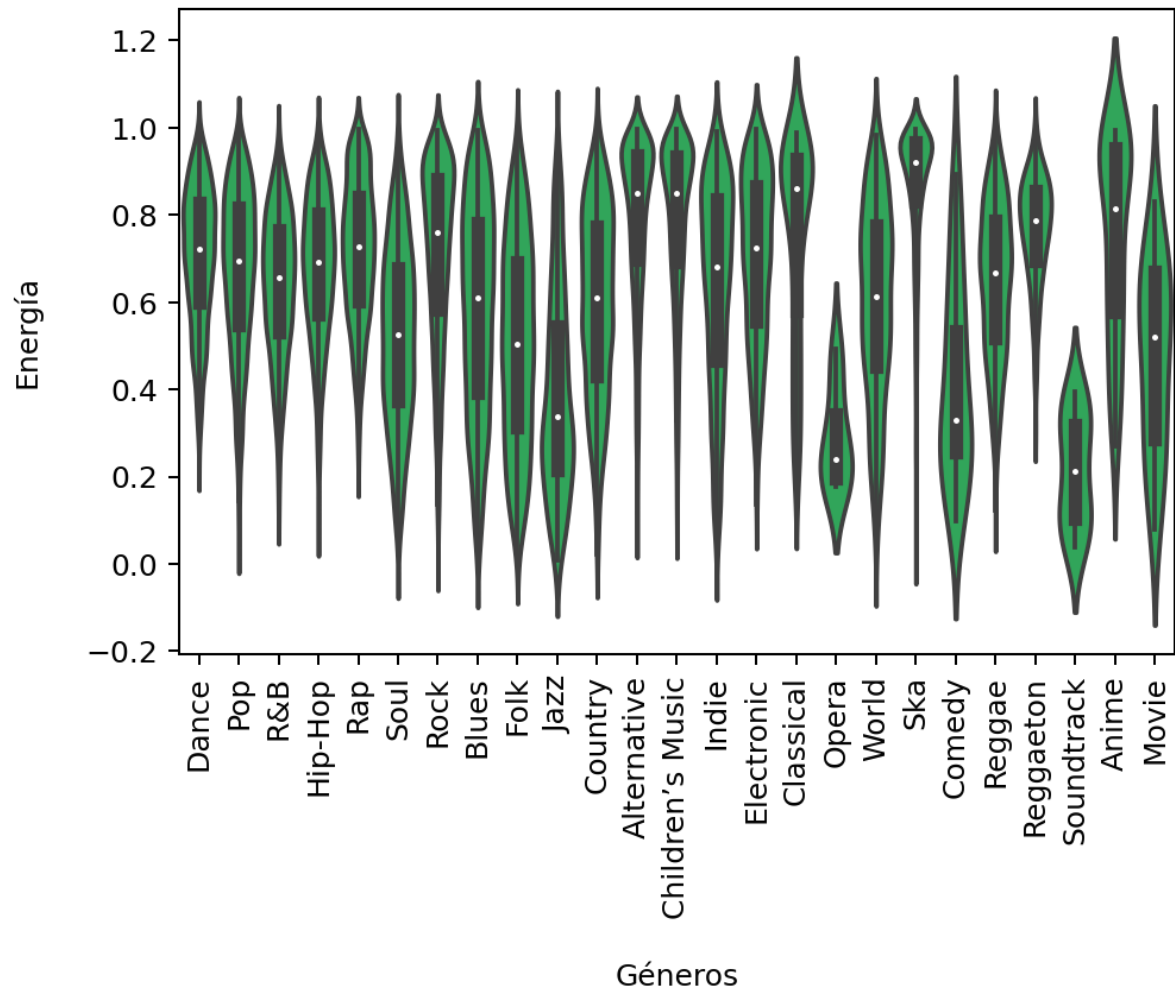


Visu 5



Visu 6

Energía de los géneros



Parte 2

Link al notebook de la parte 2:

https://colab.research.google.com/drive/14iqKYK_vckQuPYOyUFWUnlquh7XutVhZ?usp=sharing

¿Cuál es el mejor score de validación obtenido?

El mejor score de validación obtenido es 0.5501582278481013

Al predecir con este modelo para test, ¿Cuál es el score obtenido?

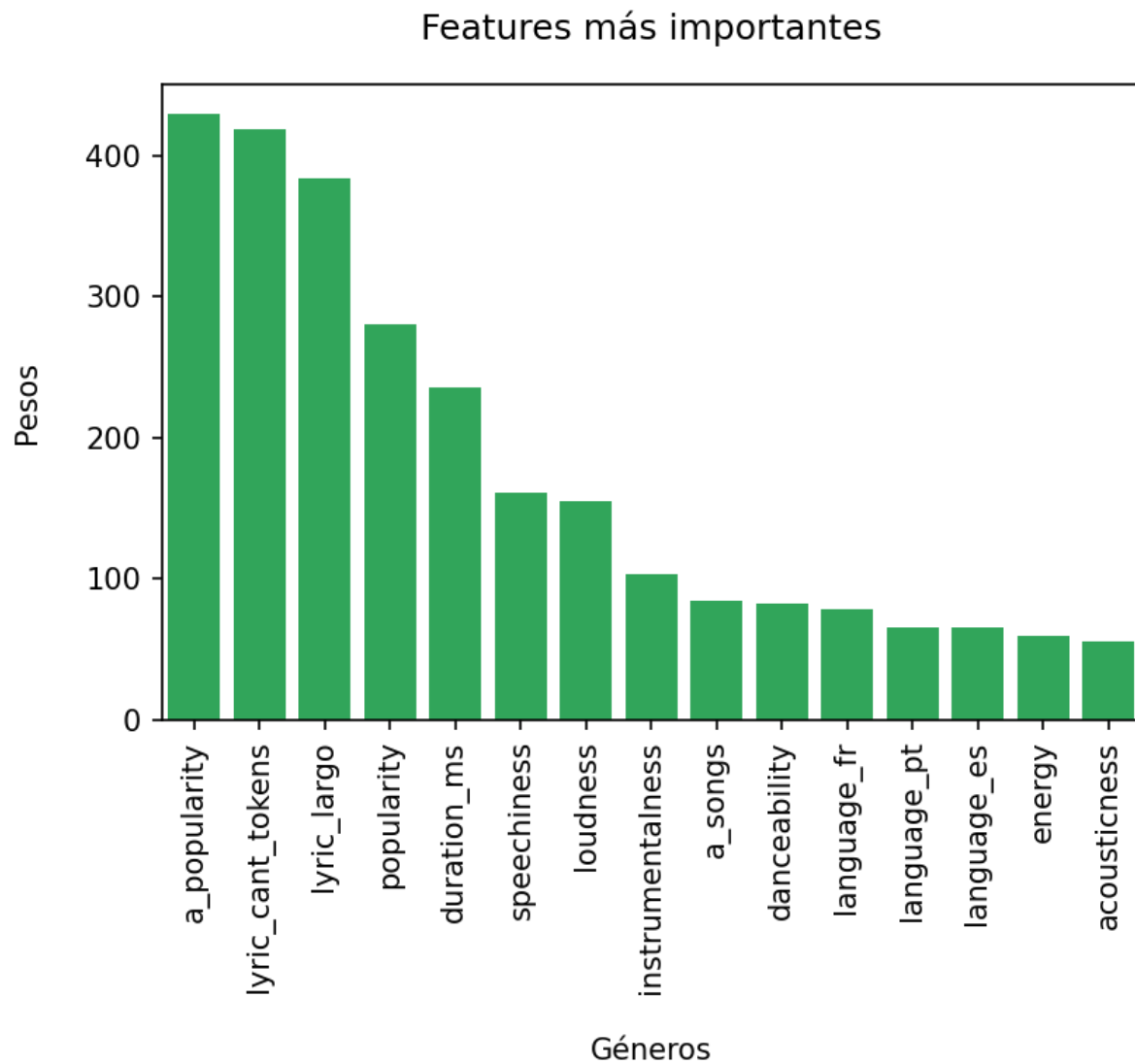
Score obtenido para test: 0.535883547731889

Link al CSV con las predicciones:

<https://drive.google.com/file/d/1fhZHFSDK3kT1hWoD-lpt6hgsTh-vHKjP/view?usp=sharing>

¿Qué features son los más importantes para predecir con el mejor modelo? Graficar.

Las features más importantes para predecir con el mejor modelo son las siguientes:



Parte III

Notebook Modelo 1 - Random Forest:

<https://colab.research.google.com/drive/1afeRPULMeS4p2GlpKAMmHOVeI9A2FIRx?usp=sharing>

Notebook Modelo 2 - XGBoost:

https://colab.research.google.com/drive/1_HQ6-OBHyWZje1v9DPFbbGu9gYDGUX-o?usp=sharing

¿Cuál es el mejor de los dos?

El mejor modelo de los dos es el de Random Forest, con un score en test de 0.529113067027759

Link al CSV con las predicciones:

https://drive.google.com/file/d/10FQzupCCxaLMxyUVbRnqxmZvO-9_V9VI/view?usp=sharing

Puntos extra (si hicieron alguno)

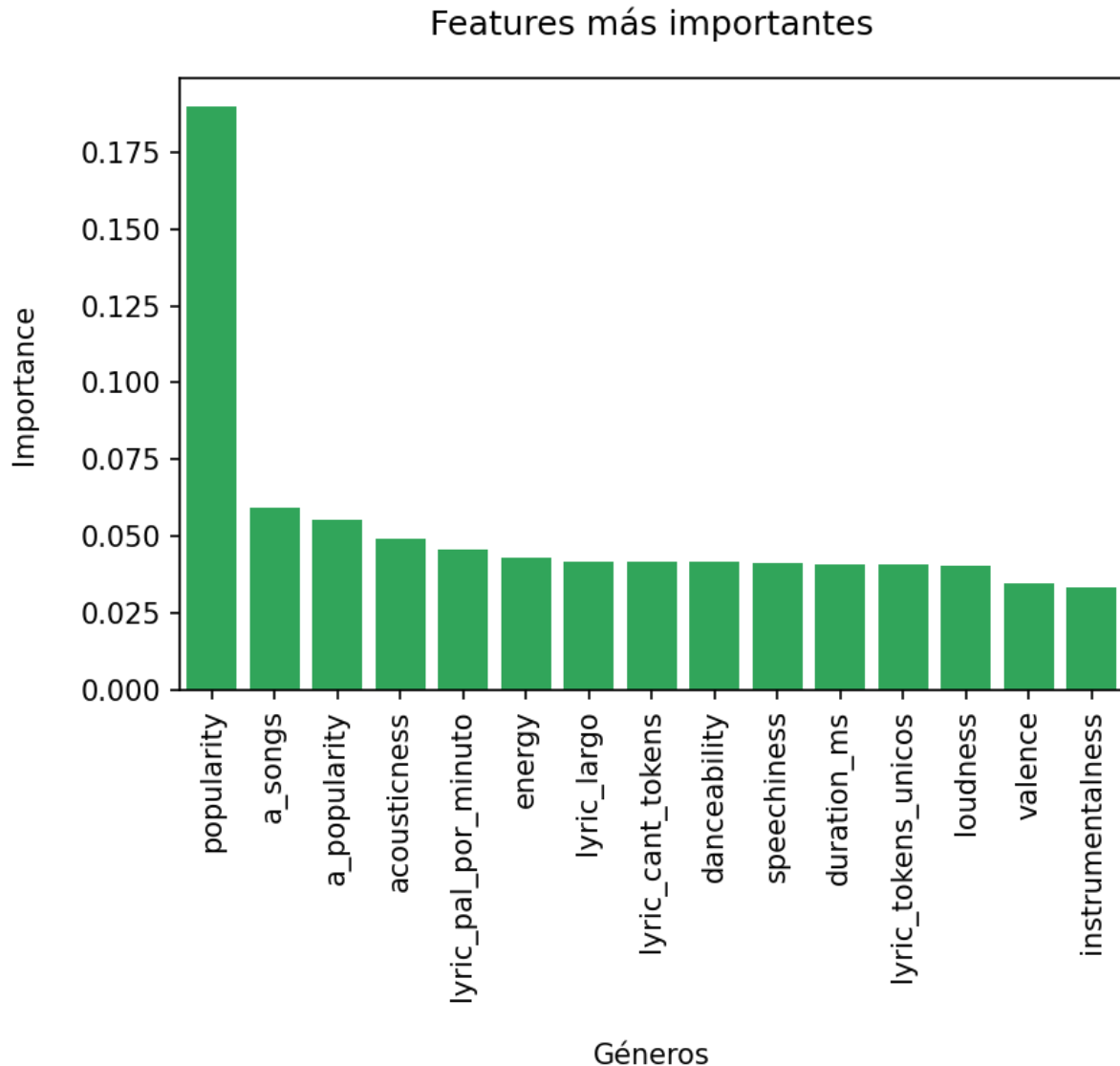
Entrenar un modelo basado en árboles que sea multi-label. ¿Qué tan bien dan las métricas en validación y test? (1 punto)

<link al notebook>

Entrenar una red neuronal con Keras que sea reproducible, usando al menos 40 features y un score en validación superior a 0,3. Debe ser un modelo por separado a los propuestos, no necesita búsqueda de hiper parámetros ni cumplir otra condición. ¿Cuál es su score en validación y en test? (1 puntos)

<link al notebook>

Graficar la importancia de features para algún modelo de la parte III.
¿Qué tanto se parece a los features importantes de la parte II? (1 punto)
<https://colab.research.google.com/drive/1afeRPULMeS4p2GlpKAMmHOVeI9A2FIRx?usp=sparing>



Vemos que las features más importantes se parecen bastante a la de la parte 1, pero con algunas diferencias. En este modelo popularity tiene mucha más importancia que todas las demás features, y no se tiene muy en cuenta a los idiomas, a diferencia de Logistic Regression. Además, aparecen aquí algunas de las features nuevas que no estaban en la parte 2.

Ensamble ambos modelos de la parte III en uno solo. ¿Cuál es su score en validación y en test? (1 punto)

<https://colab.research.google.com/drive/1afeRPULMeS4p2GlpKAMmHOVeI9A2FIRx?usp=sparing>