



Spotify Charts (All Audio Data)

CC5205 - Minería de Datos

Grupo 13: Insight Alchemists
Diego Opazo D.
Diego Soto C.
Francisco Gutiérrez A.
Nadia Decar D.
Vicente Soto

Motivación

La variedad de géneros y tipos de música y la importancia de esta en el día a día de las personas, derivan en una gran cantidad de información que puede ser estudiada de distintas formas y distintos puntos de vista.

Lo anterior, más la idea de aprender de las características que permiten identificar éxitos en el ámbito de la música nos motivan a querer estudiar este dataset

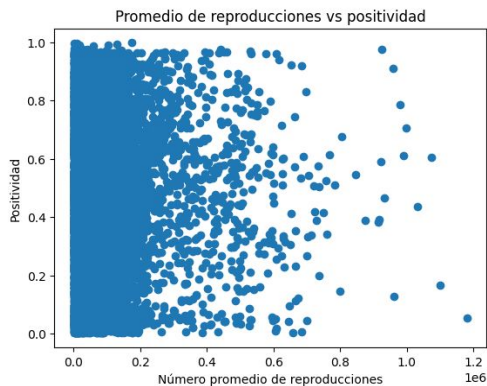
Preguntas a responder con nuestro proyecto:

- ¿Es posible predecir la felicidad o positividad de una canción en función de la popularidad u otros parámetros?
- ¿Es posible asociar el valor de energía u otros parámetros de una canción, con la región geográfica en donde es más popular?
- ¿Qué parámetros son más adecuados para predecir la popularidad de una canción?

Exploración de datos

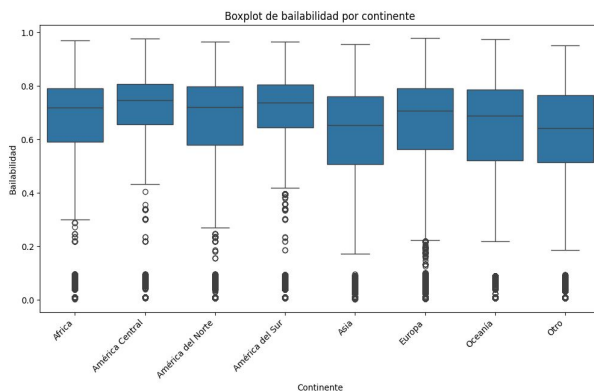
Promedio reproducciones v/s positividad:

Se exploró la relación entre la cantidad de reproducciones que tiene una canción respecto a la positividad que esta posee. Esta última se mide desde 1 es muy alegre y 0 es triste.



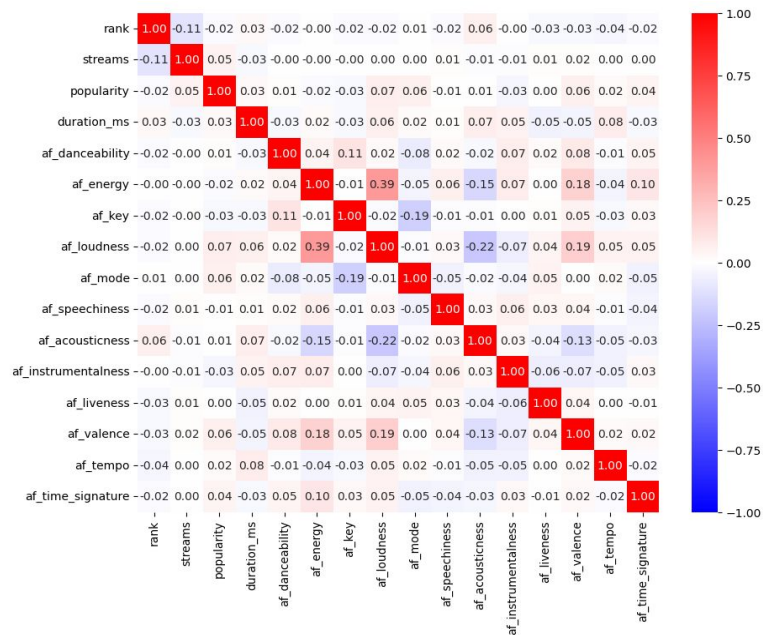
Bailabilidad:

Se evaluó la idoneidad para bailar considerando ritmo, estabilidad y beats en el top 50 de distintas regiones. De esta forma podemos ver cómo influye este valor en la pertenencia de la canción a este ranking por región.



Correlaciones:

Una parte importante al momento de realizar los experimentos, es revisar cómo se comportan los atributos entre ellos.



¿Es posible predecir la positividad de una canción basándose en su popularidad y tempo?

- Selección de parámetros útiles para la predicción.
 - Tras diversas pruebas se decide tomar: “streams”, “popularity” y “af_tempo”
- Preprocesamiento:
 - Función de etiquetado para “af_valence” (Asigna strings)
 - Escalado de los datos para normalización

- GridSearchCV:

- Optimización de parámetros

- Entrenamiento

- K-Nearest neighbors

- Resultados

	precision	recall	f1-score	support
High	0.96	0.96	0.96	58196
Low	0.95	0.95	0.95	52533
Medium-High	0.96	0.96	0.96	109647
Medium-Low	0.96	0.95	0.95	82345
accuracy			0.96	302715
macro avg	0.96	0.96	0.96	302715
weighted avg	0.96	0.96	0.96	302715

¿Es posible asociar el valor de alguna característica de la música como su energía o su tempo con un lugar geográfico?

- Modificación de parámetros para la predicción.
 - La columna de los países es modificada para reducir la cantidad de clases agrupando los países por regiones geográficas.
- Selección de una muestra del dataset.
 - Se seleccionan solo los datos que tienen que ver con las características medida de manera numérica.

• Pipeline

- Estandarización de datos numéricos.
- Separación de datos en test, val y train

• Entrenamiento

- Se usa Decision Tree Classifier

• Resultados

	precision	recall	f1-score	support
Africa	0.94	0.46	0.62	1785
America_latina	0.82	0.75	0.78	15151
America_norte	0.64	0.17	0.27	1785
Asia	0.88	0.67	0.76	8922
Europa	0.68	0.92	0.78	25448
Oceania	0.83	0.22	0.35	1785
Oriente_medio	0.87	0.38	0.53	4461
accuracy			0.74	59337
macro avg	0.81	0.51	0.59	59337
weighted avg	0.77	0.74	0.73	59337

¿Cuál será la popularidad de una canción en una región debido a sus atributos?

- Selección de parámetros útiles para la predicción.
 - Descartados: url, fechas, nombre del álbum, mercado, track_id, entre otras.
- Selección de una muestra del dataset.
- Pipeline
 - Preprocesamiento de texto.
 - Estandarización de datos numéricos.
 - Random Forest.

• Entrenamiento

• Resultados

	precision	recall	f1-score	support
High	0.93	1.00	0.97	115
Low	0.77	0.74	0.75	54
Medium-High	0.94	0.90	0.92	67
Medium-Low	0.75	0.72	0.74	64
accuracy			0.87	300
macro avg	0.85	0.84	0.84	300
weighted avg	0.87	0.87	0.87	300



Spotify Charts (All Audio Data)

CC5205 - Minería de Datos

Grupo 13: Insight Alchemists
Diego Opazo D.
Diego Soto C.
Francisco Gutiérrez A.
Nadia Decar D.
Vicente Soto