



Recomendador musical contextualizado (PIPELINE)

QUICAHÑO MIRANDA, Victor Alejandro

Docente: Prof Mag.VALDIVIA CUADROS, Ana María
Curso: Tópicos en Ciencia de Datos

UNSA - Universidad Nacional de San Agustín de Arequipa
Junio de 2025

Abreviaturas

Índice

1. Descripción del Dataset	6
1.1. Music4All [1]	6
1.1.1. Contexto	6
1.1.2. Entidad de estudio	7
1.2. Music4All - Onion [2]	7
1.2.1. Contexto	7
1.2.2. Entidad de estudio	7
1.3. Extrasensory [3]	8
1.3.1. Contexto	8
1.3.2. Entidad de estudio	8
2. Cuatro Preguntas	9
2.1. ¿Qué problemas identificas en el dataset?	9
2.2. ¿Qué descubriste al analizar los datos?	9
2.3. Qué reflejan los patrones de tendencia?	10
2.4. ¿Cómo es afectado el comportamiento humano (registros musicales) en relación con el tiempo?	11

Índice de cuadros

1.	Descripción de atributos del conjunto de datos musicales	15
2.	Descripción de archivos de características musicales	16
3.	Descripción de los archivos y atributos utilizados en el análisis	17
4.	Archivos relacionados con metadatos generados por usuarios	17
5.	Archivos con representaciones visuales extraídas de videos musicales	18
6.	Resumen de archivos de historial de escucha	18
7.	Descripción de modalidades, cantidad de variables y características del dataset	19

Índice de figuras

1.	Pipeline de la elaboración de Music4All.	6
2.	Capas de Music4All-Onion.	8
3.	Histogr�ma de datos faltantes por mes.	10
4.	Matriz de correlaci�n entre actividad y energia promedio.	11
5.	Nube de palabras por los g�neros m�s �idos por los 33 usuarios con m�s registros.	12
6.	Histograma de los 20 g�neros musicales m�s �idos por los 33 usuarios con m�s registros.	13
7.	Mapa de calor de la cantidad de canciones distintas escuchadas por a�o, para los 33 usuarios con m�s registros.	14
8.	Histograma que muestra la cantidad de “nuevas canciones” escuchadas cada a�o por los 33 usuarios con mayores registros.	14

1. Descripción del Dataset

1.1. Music4All [1]

1.1.1. Contexto

El dataset Music4All surge como una contribución a la comunidad de Music Information Retrieval (MIR), cuyo objetivo principal es el desarrollo de sistemas que permitan recuperar y recomendar contenido musical de manera eficiente y efectiva. Pese a los avances en el área, existía una carencia de bases de datos que cumplieran con ciertos requisitos fundamentales para la investigación:

- Disponibilidad de señales de audio
- Diversidad de atributos musicales
- Un volumen considerable de piezas musicales.

Es así que surge el DataSet Music4All, desarrollado para cubrir dichas necesidades. Dentro de sus bondades ofrece un conjunto de datos amplio y diverso que incluye información de más de **109,000 canciones** y el historial de escucha de **15,602 usuarios anónimos**. El dataset contiene diversos tipos de información para cada canción, incluyendo metadatos, clips de audio de 30 segundos, letras, información de popularidad y atributos acústicos derivados de la API de Spotify, así como etiquetas de género y tags generados por usuarios en Last.fm.



Figura 1: Pipeline de la elaboración de Music4All.

1.1.2. Entidad de estudio

La entidad principal de estudio en Music4All es la canción. Cada canción está descrita por una serie de atributos que pueden agruparse en categorías como metadatos, características acústicas, y etiquetas semánticas [**Revisar la estructura del dataset en el Cuadro: 1**]

1.2. Music4All - Onion [2]

1.2.1. Contexto

El dataset Music4All-Onion es una extensión multimodal y a gran escala del dataset Music4All. Este fue diseñado para suplir las limitaciones comunes en los sistemas de recomendación musical actuales: la dependencia exclusiva de interacciones usuario-item (colaborative filtering), ignorando el contenido musical como base para la afinidad del usuario. Esta omisión se debe principalmente a la falta de datasets estandarizados que integren tanto información colaborativa como de contenido.

Para solucionar este problema, Music4All-Onion incorpora 26 nuevas características de tipo audio, video y metadatos, además de proporcionar 252,984,396 registros de escucha pertenecientes a 119,140 usuarios, extraídos de Last.fm. Esto permite la investigación comparativa del impacto de distintas capas semánticas del contenido musical (modelo tipo cebolla”) sobre la precisión, novedad y equidad de los sistemas de recomendación.

1.2.2. Entidad de estudio

Music4All-Onion organiza la información del contenido musical en capas (layers), que representan distintos tipos de características semánticas. La entidad de estudio principal continúa siendo la canción, pero se analiza desde múltiples perspectivas (audio, texto, imagen, interacción):

1. **Capa de Audio (Audio Layer)** Contiene descriptores extraídos de la señal de audio con herramientas como OpenSMILE y Essentia. [**Revisar el Cuadro: 2**]
2. Capa de Contenido Derivado - Texto (Extracted Metadata/Derived, EMD) [**Revisar el Cuadro: 3**]
3. Capa de Contenido Generado por el Usuario (User-Generated Content, UGC) [**Revisar el Cuadro: 4**]
4. Capa de Contenido Derivado Visual (Derived Content, DC) [**Revisar el Cuadro: 5**]
5. Eventos de Escucha (Interacciones Usuario-Canción) [**Revisar el Cuadro: 6**]

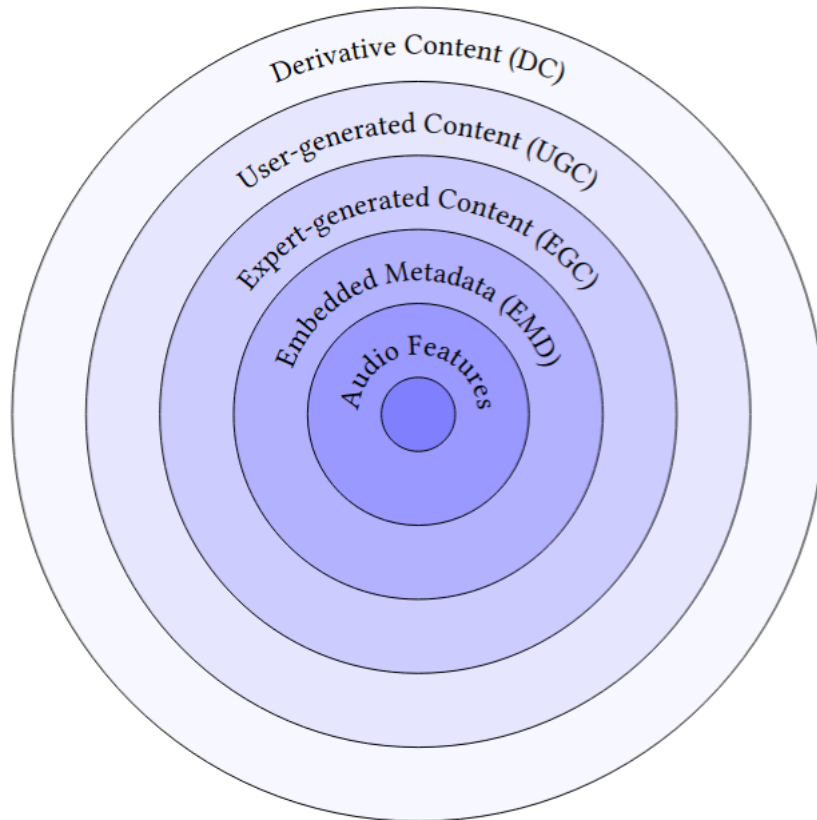


Figura 2: Capas de Music4All-Onion.

1.3. Extrasensory [3]

1.3.1. Contexto

El dataset ExtraSensory fue recopilado “en la vida real” mediante la aplicación móvil ExtraSensoryApp desarrollada en la UC San Diego. Durante aproximadamente un año (2015–2016), 60 voluntarios llevaron su smartphone y en algunos casos, también un smartwatch configurados para registrar automáticamente muestras de datos sensoriales cada minuto. Al mismo tiempo, los usuarios auto-reportaban su actividad y su contexto (por ejemplo, “caminando”, “sentado”, “en reunión”, “en casa”) mediante una interfaz de la app. Este diseño permite estudiar el reconocimiento de contexto humano y la actividad diaria bajo condiciones naturales, superando las limitaciones de los experimentos controlados en laboratorio

1.3.2. Entidad de estudio

La entidad de análisis es una **muestra de contexto**, dentro de un vector de características sensoriales y de estado asociado a un instante (1 min) y al usuario que lo generó. Cada muestra incluye: **[Revisar el Cuadro 7 para conocer la estructura del Dataset]**

- Metadatos de identificación
 - uuid – Identificador único de usuario (60 IDs)
 - timestamp – Marca temporal de la muestra
- Etiquetas de contexto
 - 50 columnas binarias, una por cada etiqueta (actividad o situación), con valor 1 si aplica, 0 si no.
 - Ejemplos: label_walking, label_sitting, label_meeting, label_home, etc
- Características sensoriales y de estado
 - Un total de 225 features agrupadas en seis grandes modalidades: acelerómetro, giroscopio, magnetómetro, audio, estado del teléfono y brújula del smartwatch.
 - Para cada sensor se extraen estadísticas temporales (media, mediana, desviación estándar, percentiles, ...)

2. Cuatro Preguntas

2.1. ¿Qué problemas identificas en el dataset?

En el conjunto de datos Extrasensory la anotación de actividades y contextos dependió en gran medida de que los propios participantes completaran manualmente los campos de labels; por ello aparecen intervalos sin etiquetas. Además, algunos usuarios tuvieron dificultades para sincronizar el teléfono con el smartwatch y, en ocasiones, la señal de sensores como GPS o acelerómetro se interrumpió, lo que originó vacíos adicionales. Sin embargo, todas estas variables están registradas como series temporales con marcas de tiempo uniformes, lo que permite aprovechar la continuidad temporal: mediante Quadratic Interpolation (QI)[4] podemos estimar los valores faltantes ajustando un polinomio de segundo grado entre las observaciones vecinas y calculando puntos intermedios suaves y coherentes. Este enfoque conserva la tendencia local y las aceleraciones propias de la señal mejor que la interpolación lineal, reduciendo el sesgo introducido por las ausencias y manteniendo la fidelidad de las secuencias originales para los análisis posteriores.

2.2. ¿Qué descubriste al analizar los datos?

Tras un análisis exhaustivo de los registros mensuales, se identificó una relación significativa entre las actividades realizadas por los usuarios y la energía promedio de las canciones reproducidas. El estudio reveló los patrones de consumo musical a lo largo de un mes, la energía musical (medida en escala normalizada 0.0-1.0 según los descriptores de

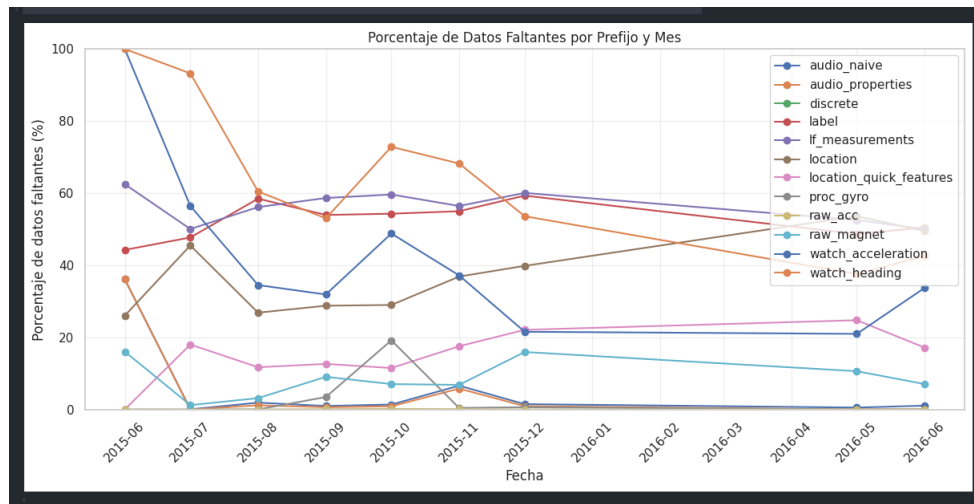


Figura 3: Histograma de datos faltantes por mes.

Spotify) mostró variaciones consistentes en correlación. siendo actividades con un mayor a fin aquellas que poseían una energía promedio mayor a comparación de otras un tanto más monótonas

2.3. Qué reflejan los patrones de tendencia?

En las Figuras 5 y 6 se presenta la distribución de géneros musicales más escuchados por los 33 usuarios con mayor número de registros de escucha.

La **nube de palabras** (Figura 5) permite observar de forma visual e intuitiva cuáles son los géneros más predominantes en el conjunto de datos. Tal como se aprecia, los géneros **pop** y **rock** destacan significativamente por encima del resto, lo cual podría introducir un sesgo importante en el análisis y en la construcción de modelos de recomendación. Esta sobrerepresentación puede limitar la capacidad del sistema para generalizar adecuadamente hacia géneros menos frecuentes.

Por otro lado, el **histograma de frecuencias** (Figura 6) brinda una visión cuantitativa más detallada de los 20 géneros más escuchados. Si bien pop y rock lideran con gran diferencia, el resto de los géneros muestra una distribución más uniforme, lo cual abre la posibilidad de segmentar o agrupar recomendaciones según preferencias más específicas.

Es importante destacar que **una misma canción puede estar etiquetada con múltiples géneros**, lo cual añade riqueza semántica y permite capturar matices más finos en las preferencias musicales. Esta característica puede aprovecharse para mejorar los sistemas de recomendación a través del análisis multilabel o estrategias de representación de género más complejas.

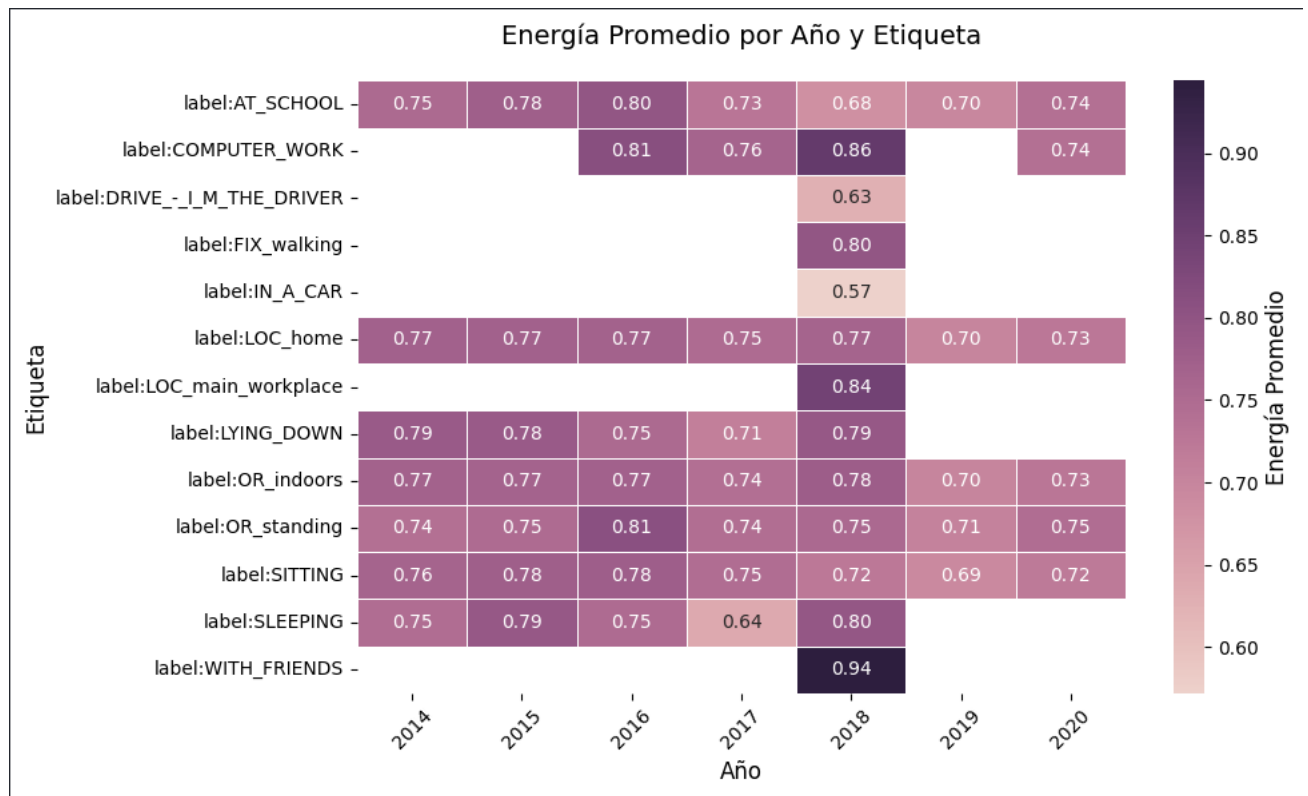


Figura 4: Matriz de correlación entre actividad y energía promedio.

2.4. ¿Cómo es afectado el comportamiento humano (registros musicales) en relación con el tiempo?

Los gráficos presentados en las Figuras 7 y 8 permiten observar cómo varían los registros musicales de los usuarios a lo largo del tiempo:

En el **mapa de calor** (Figura 7), cada fila representa un usuario y cada columna, un año calendario. Se aprecia que la mayoría de usuarios no exhibe una tendencia clara a incorporar nuevos temas año tras año; los espacios correspondientes a “canciones nuevas” suelen estar prácticamente vacíos o con muy pocos eventos. Este comportamiento puede explicarse, en parte, por la naturaleza de la población analizada—procedente del estado de California—donde los oyentes tienden a mantener listas de reproducción estables y consumir principalmente éxitos consolidados. De hecho, estudios basados en datos de YouTube Music indican que los temas más reproducidos en California se mantienen casi invariables durante periodos prolongados, sin que la mayoría del público incorpore constantemente lanzamientos recientes :contentReference[oaicite:0]index=0.

En el **histograma** (Figura 8), que muestra la frecuencia de “nuevas canciones escuchadas” por año para los 33 usuarios con más registros, se confirma que el recuento de temas inéditos permanece bajo incluso en años con alta actividad de escucha. Esto sugiere que, aun cuando un usuario sea muy activo, su repertorio no se actualiza significativamente

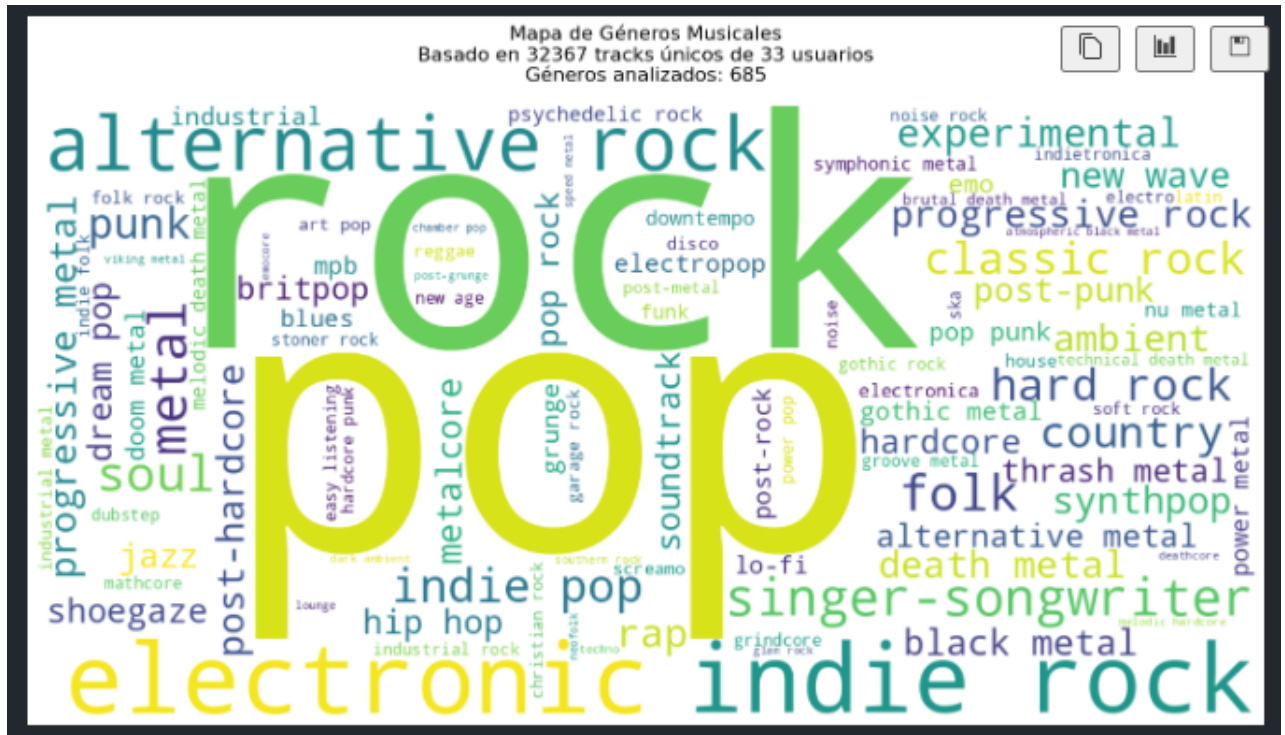


Figura 5: Nube de palabras por los géneros más oídos por los 33 usuarios con más registros.

cada temporada. Por ende, el análisis temporal revela dos aspectos clave:

- Los gustos musicales de estos usuarios son *estables* a lo largo del tiempo, con poca rotación de repertorio.
- Existen periodos de mayor o menor actividad según la energía de las canciones reproducidas (ver variación en la Figura 8), pero esto no implica necesariamente la incorporación de nuevos temas.

En conjunto, estas observaciones ponen de manifiesto que la dimensión temporal y la falta de novedad en el repertorio son factores relevantes para entender y modelar el comportamiento de escucha en un entorno *context-aware*.

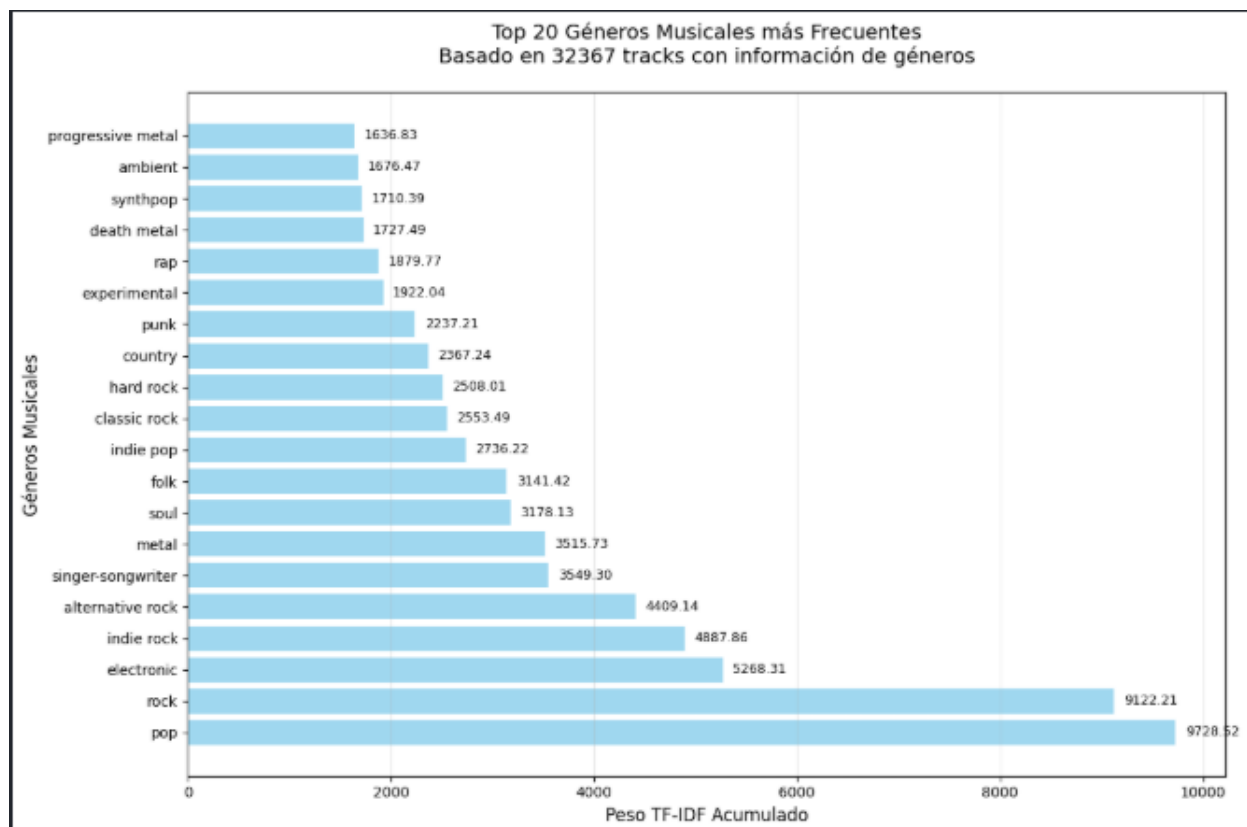


Figura 6: Histograma de los 20 géneros musicales más oídos por los 33 usuarios con más registros.

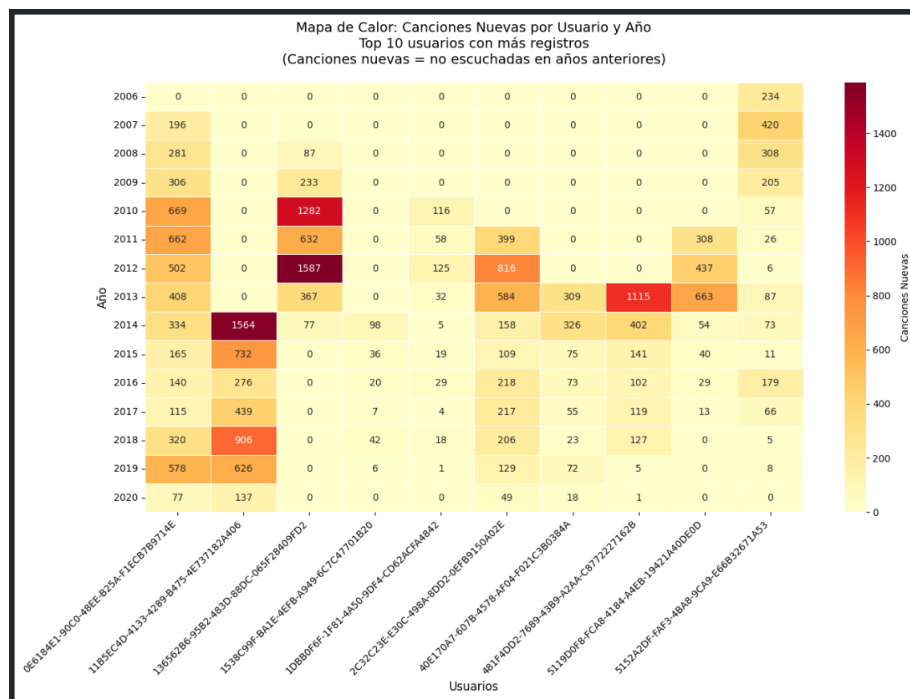


Figura 7: Mapa de calor de la cantidad de canciones distintas escuchadas por año, para los 33 usuarios con más registros.

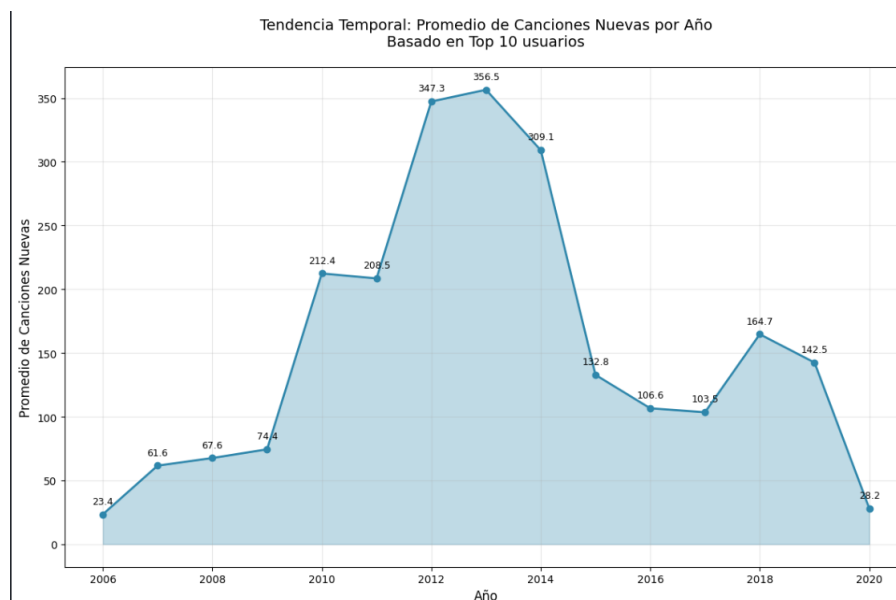


Figura 8: Histograma que muestra la cantidad de “nuevas canciones” escuchadas cada año por los 33 usuarios con mayores registros.

Cuadro 1: Descripción de atributos del conjunto de datos musicales

Atributo	Descripción	Tipo de dato	Rango/Valores posibles
id	Identificador único de la canción (16 caracteres)	Texto (string)	–
artist	Nombre del artista según Last.fm	Texto (string)	16,269 artistas únicos
song	Título de la canción	Texto (string)	–
lang	Idioma detectado de la letra	Texto (string)	46 idiomas únicos
spotifyid	Identificador de la canción en Spotify	Texto (string)	–
popularity	Nivel de popularidad en Spotify	Entero	0 a 100
album name	Nombre del álbum	Texto (string)	38,363 álbumes únicos
release	Año de publicación	Entero	Año (ej. 1960–2020)
danceability	Medida de qué tanailable es una canción	Decimal (float)	0.0 a 1.0
energy	Intensidad o actividad percibida en la canción	Decimal (float)	0.0 a 1.0
key	Tónica o clave principal (Pitch Class notation)	Entero	0 (C) a 11 (B)
mode	Modalidad (mayor o menor)	Binario (entero)	0 = menor, 1 = mayor
valence	Medida de positividad o felicidad	Decimal (float)	0.0 a 1.0
tempo	Tempo musical en beats por minuto	Decimal (float)	Depende de la canción (ej. 60–200 BPM)
genres	Lista de etiquetas de género asociadas a la canción	Lista de strings	853 géneros únicos
tags	Etiquetas generadas por usuarios en Last.fm	Lista de strings	19,541 etiquetas únicas

Cuadro 2: Descripción de archivos de características musicales

Archivo	Descripción técnica	Tipo
id_blf_*	Patrones espectrales (Band Energy, Spectral Flux) y patrones de fluctuación (BLF - Band-wise Local Feature) para análisis temporal-espectral	Audio BLF
id_chroma_bow.tsv.bz2	Representación Bag-of-Words (BoW) de características cromáticas (Chroma) que capturan la información tonal en 12 clases de altura	Audio OpenSMILE
id_compare_*.tsv.bz2	Estadísticas acústicas del conjunto ComPa-rE: MFCCs (Mel-Frequency Cepstral Coefficients), pitch fundamental, shimmer, jitter y parámetros espectrales	Audio OpenSMILE
id_emobase_*.tsv.bz2	Características acústicas relacionadas con emociones (EmoBase), incluyendo prosodia, calidad vocal y descriptores espectrales emocionales	Audio OpenSMILE
id_essentia.tsv.bz2	Descriptores musicales avanzados calculados con Essentia: características espectrales (centroide, flatness), rítmicas (BPM, onset rate) y tonales (key, scale)	Audio Essentia
id_mfcc_*.tsv.bz2	Coeficientes Cepstrales de Frecuencia Mel (MFCC) con sus estadísticas y representación Bag-of-Audio-Words (BoAW) para modelado tímbrico	Audio MFCC
id_ivec*.tsv.bz2	I-vectors (vectores de identidad) generados mediante Modelos de Mezcla Gaussiana (GMM) con diferentes configuraciones (256, 512 y 1024 componentes)	Audio MFCC

Atributo / Archivo	¿Qué representa?	Cómo se obtiene
processed_lyrics	Letras preprocesadas: normalizadas, lematizadas, sin anotaciones musicales	Procesamiento lingüístico y limpieza
id_lyrics_sentiment_functionals	Valores de valencia, activación y dominancia (emociones)	Léxico afectivo ANEW
id_lyrics_tf-idf	Representación TF-IDF (qué tan relevantes son las palabras en el corpus total)	Modelado de texto
id_lyrics_word2vec	Embeddings vectoriales de palabras, promediados por canción	Modelo Word2Vec preentrenado
id_vad_bow	Medidas emocionales según VADER, en formato "bag of words"	Léxico de sentimiento VADER

Cuadro 3: Descripción de los archivos y atributos utilizados en el análisis

Archivo	Descripción
id_genres_tf-idf.tsv	Representación TF-IDF de géneros musicales
id_tags_dict.tsv	Tags generados por usuarios de Last.fm con pesos asociados
id_tags_tf-idf.tsv2.bz2	Representación TF-IDF de tags

Cuadro 4: Archivos relacionados con metadatos generados por usuarios

Archivo	Descripción	Tipo (DC)
id_incp.tsv.bz2	Vectores de imágenes extraídas de videos usando Inception v3	Imagen
id_resnet.tsv.bz2	Vectores de imágenes extraídas con ResNet	Imagen
id_vgg19.tsv.bz2	Vectores de imágenes extraídas con VGG19	Imagen

Cuadro 5: Archivos con representaciones visuales extraídas de videos musicales

Archivo	userid_trackid_timestamp.tsv	userid_trackid_count.tsv
Descripción	Eventos individuales de escucha con timestamp	Conteo de reproducciones por usuario por canción
Nº Usuarios	119,140	119,140
Nº Canciones	56,512	56,512
Nº Registros	252,984,396	50,016,042

Cuadro 6: Resumen de archivos de historial de escucha

Modalidad	Cantidad	Descripción	Tipo de dato	Rango / Valores
Acelerómetro	26	Estadísticos (mean, std, median, ...) de aceleración en ejes X/Y/Z	Decimal (float)	Física (m/s^2)
Giroscopio	26	Estadísticos de velocidad angular en ejes X/Y/Z	Decimal (float)	Física ($^\circ/\text{s}$)
Magnetómetro	26	Estadísticos de campo magnético en ejes X/Y/Z	Decimal (float)	Física (μT)
Audio	13	Estadísticos de nivel sonoro (RMS, peak, centroid, ...)	Decimal (float)	Intensidad (varía según clip)
Estado teléfono	7	Indicadores de estado: nivel batería (0–1), Wi-Fi (on/off), Bluetooth (on/off), modo avión, etc.	Mix: float, binario	batería 0–1, flags 0/1
Brújula watch	6	Estadísticos de orientación (mean, std) extraídos del compass del smartwatch	Decimal (float)	Grados (0–360)
Etiquetas	~50	Self-reports binarios de actividades y situaciones	Binario (0/1)	0 = no aplica, 1 = aplica
Metadatos	2	uuid, timestamp	String, datetime	—

Cuadro 7: Descripción de modalidades, cantidad de variables y características del dataset

Referencias

- [1] I. A. Pegoraro Santana, F. Pinhelli, J. Donini, L. Catharin, R. B. Mangolin, Y. M. e. G. da Costa, V. Delisandra Feltrim, and M. A. Domingues, “Music4all: A new music database and its applications,” in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, Jul. 2020, p. 399–404. [Online]. Available: <http://dx.doi.org/10.1109/IWSSIP48289.2020.9145170>
- [2] M. Moscati, E. Parada-Cabaleiro, Y. Deldjoo, E. Zangerle, and M. Schedl, “Music4all-onion – a large-scale multi-faceted content-centric music recommendation dataset,” in *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, ser. CIKM '22. ACM, Oct. 2022, p. 4339–4343. [Online]. Available: <http://dx.doi.org/10.1145/3511808.3557656>
- [3] Y. Vaizman, K. Ellis, G. Lanckriet, and N. Weibel, “ExtraSensory app,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, Apr. 2018, pp. 1–12.
- [4] Anónimo, Mar. 2024. [Online]. Available: <https://newdigitals.org/2024/03/14/time-series-imputation-interpolation-anomaly-detection/>