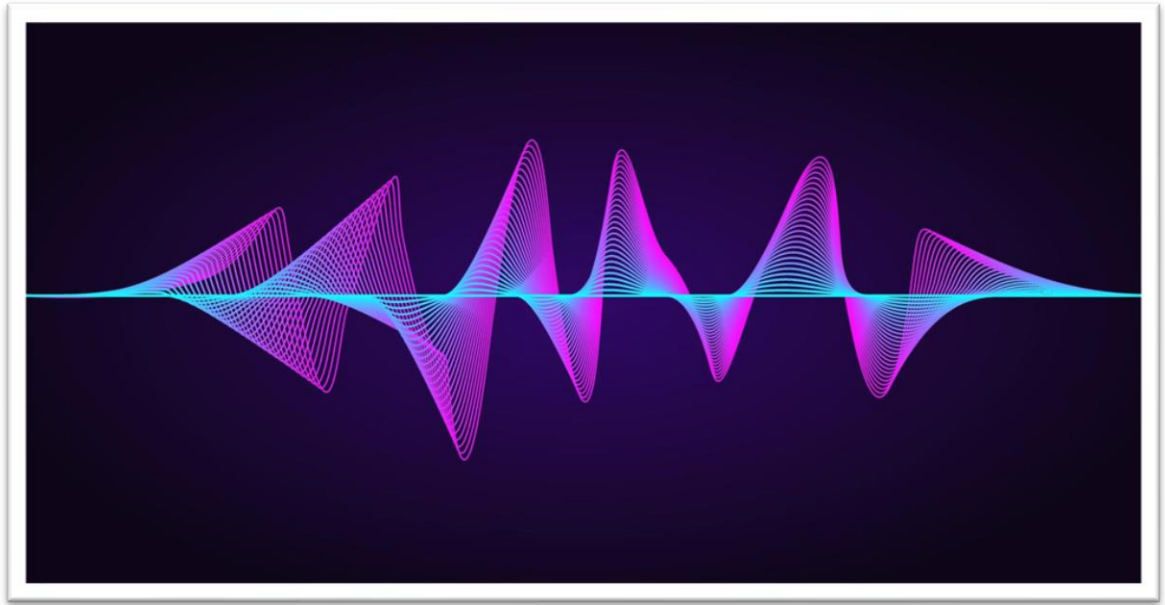


DISTRIBUCIÓN DIGITAL DE LA MÚSICA

EXPLORATORY DATA ANALYSIS (EDA)



Sebastián Pomi Rodríguez

Agosto 2023

RESUMEN

Este trabajo de investigación se enfoca en el análisis de datos de la distribución digital de música a través de plataformas como Spotify y Youtube. Se utiliza un Dataset con más de veinte mil canciones de diversos artistas para identificar patrones de popularidad y su relación con características musicales. Se formulan dos hipótesis: que las canciones populares son más bailables y alegres, y que las canciones más escuchadas tienen una duración corta. Los resultados indican que las canciones populares tienden a ser más bailables, más alegres y menos enérgicas, aunque las diferencias no son significativas, y que las canciones con duración corta son las más escuchadas. El análisis proporciona una visión detallada sobre cómo el público interactúa con la música en el entorno digital y cómo las características musicales influyen en la preferencia de las canciones. Estos hallazgos ofrecen conocimientos valiosos para comprender la industria musical actual y sus dinámicas en la era digital.

INDICES

RESUMEN.....	2
INTRODUCCIÓN	5
Hipótesis:.....	6
Otras consultas	6
DESARROLLO DEL CONTENIDO.....	7
OBTENCIÓN DE DATOS:	7
TRATAMIENTO DE COLUMNAS.	7
Traducción de columnas:.....	8
Conversión de columnas:	8
Creación de columnas:	8
Tratamiento de datos:	8
TRATAMIENTO DE DATOS NULOS	8
LIMPIEZA DE DATOS POR VARIABLE	9
Hablabilidad:	9
Canción, Artista y Álbum.	9
Instrumentalidad:.....	9
ANÁLISIS DE DATOS	10
ANÁLISIS UNIDIMENSIONAL	10
ANÁLISIS MULTIDIMENSIONAL.....	10
Top Artistas	10
Top Canciones	11
Top likes	12
Top comentarios	12
Análisis de correlación	13
Hipótesis 1:.....	16
Las canciones más populares son las más bailables, alegres y enérgicas.....	16
Prueba de Mann-Withney.....	16
Conclusión Hipótesis 1:	18
Hipótesis 2:.....	19
Las canciones más escuchadas no pueden tener una duración muy larga	19
Prueba de Kruskal-Wallis.....	19
Prueba de Mann-Withney.....	19
Conclusión Hipótesis 2:	19
REFERENCIAS	21

CONCLUSIONES.....	21
ANEXOS.....	22
REFERENCIAS DE COLUMNAS	22

INTRODUCCIÓN

¿Qué es la música?

La música es el arte de crear y combinar de manera armónica y melódica sonidos y silencios que sean agradable a los oídos. Estos sonidos son generados a través de la voz o de instrumentos musicales.

Pero la música también es una industria gigante que consiste principalmente en los siguientes campos

- 1- Creación y Composición
- 2- Producción musical
- 3- Distribución
- 4- Promoción y Marketing
- 5- Ventas y Comercialización

La distribución digital ha crecido de manera exponencial en los últimos años debido al auge de las tecnologías digitales y el acceso a internet en todo el mundo. Como principal actor en la distribución digital de música figuran las plataformas de streaming. Estas plataformas permiten a los usuarios acceder a un amplio catálogo en línea donde los usuarios pueden escuchar música on demand o disfrutar de playlist creadas por algoritmos que se actualizan constantemente. Aparte del cambio en la manera de escuchar música, esta digitalización está condicionando no solo el consumo de música, sino también su creación y composición.

El presente trabajo de investigación tiene como objetivo realizar un Análisis Exploratorio de Datos tomando como base un Dataset que contiene información sobre canciones de distintos artistas de todo el mundo, recopilado el 7 de febrero de 2023 en las plataformas musicales Youtube y Spotify. El Dataset incluye atributos, características y métricas de más de veinte mil canciones. Incluye atributos como nombre de la canción, del artista y del álbum, características musicales como la bailabilidad, la energía y la emoción, y métricas de escuchas, likes y comentarios

Nuestra investigación intenta responder preguntas como cuáles son los artistas más populares en cada plataforma, así como identificar las canciones más exitosas y analizar las interacciones del público con ellas. Además, evaluamos si existe una relación entre las métricas de una canción y sus características musicales, particularmente si las canciones más populares tienden a ser más enérgicas y bailables.

Este análisis nos permitirá obtener una visión más profunda sobre la relación entre las características musicales y la popularidad de las canciones. Además, ayudará a entender cómo el público interactúa con la música en el entorno digital y cuáles son los atributos que influyen en su preferencia.

A través de este EDA, esperamos aportar conocimientos relevantes y útiles para comprender mejor la industria musical actual y cómo los atributos musicales influyen en la percepción y elección de las canciones por parte de los oyentes en el entorno digital.

Hipótesis:

- 1- Las canciones más populares son las más bailables, alegres y enérgicas
Nos basamos en la idea de que a las personas les atraen especialmente las canciones tienen características positivas
- 2- Las canciones más escuchadas no pueden tener una duración muy larga.
Partimos de la premisa que las personas no suelen escuchar muchas veces una canción muy larga.

Estudiaremos características musicales de canciones y sus métricas para poder contestar esta pregunta

Otras consultas

- Artistas más populares en Spotify
- Artistas más populares en Youtube
- Canciones más escuchadas en ambas plataformas
- Canciones con mayor interacción (likes y comentarios)

DESARROLLO DEL CONTENIDO

OBTENCIÓN DE DATOS:

Para este análisis hemos obtenido los datos a partir de un Dataset que contiene información sobre canciones de distintos artistas de todo el mundo, recopilado el 7 de febrero de 2023. Cada artista tiene como máximo 10 canciones y en la base de datos se encuentran 2079 artistas.

El Dataset lo obtuvimos de la página Kaggle, compartido por el usuario Salvatore Rastelli (<https://www.kaggle.com/Datasets/salvatorerastelli/spotify-and-youtube>)

Información del Dataset

- Tiene 28 columnas
- Tiene 20.718 filas
- Algunas columnas tienen valores nulos
- Tiene 3 tipos de datos. Objet, Int y Float

Obtenemos los primeros datos:

Como primera visualización del Dataset realizamos estudios para ver el tamaño del Dataset, los nombres de las columnas, la cantidad de canciones y artistas, los tipos de datos, la cantidad de valores nulos por variable y la cantidad de valores que son cero por variable. También realizamos el primer resumen estadístico de las columnas numéricas para tener una visión general y rápida de las variables.

TRATAMIENTO DE COLUMNAS.

En este paso eliminamos las columnas que no nos sirven, las traducimos al español y tratamos sus valores nulos y ceros.

Eliminación de columnas

Decidimos eliminar 10 columnas del Dataset porque no proporcionaban datos relevantes para el análisis o los datos no eran de buena calidad.

Las columnas eliminadas fueron 'Unnamed:0', 'Description', 'Key', 'Licensed', 'official_video', 'Description', 'Url_spotify', 'Url_youtube', 'Uri', 'Channel'.

La columna 'Unnamed:0' la eliminamos porque se genera por error y no brinda información relacionada con el Dataset.

La columna 'Key' si bien es un dato característico de la canción, no la tendremos en cuenta para este análisis por la complejidad de los valores que representa. No será necesaria para nuestro estudio.

Las columnas 'Description', 'Licensed', 'official_video', 'Description', 'Url_spotify', 'Url_youtube', 'Uri', 'Channel'. Las eliminamos porque no brindan datos necesarios para el estudio que realizaremos.

Traducción de columnas:

Renombramos las columnas con su nombre en español para unificar el idioma del estudio.

Conversión de columnas:

Convertimos los valores de la columna 'Duracion_ms' de milisegundos a minutos y renombramos la columna a 'Duracion_min'.

Creación de columnas:

Agregamos la columna 'Canción - Artista' para agrupar en una columna la canción con su artista y sea más fácil de entender futuros gráficos

Tratamiento de datos:

Trataremos los datos nulos, los datos ceros y los outliers.

TRATAMIENTO DE DATOS NULOS

Podemos observar que hay datos nulos en cuanto a visualizaciones, likes, comentarios y reproducciones. Al faltar datos, buscamos la media de cada dato por artista y rellenamos los datos nulos con estos valores representativos.

Luego de completar con la media por artista, aún quedan valores nulos en las variables 'Likes_YT' y 'Comentarios_YT' y decidimos completar estos nulos con un cero.

También faltan datos con la columna 'Titulo_video' por lo que decidimos completar esos valores nulos con la unión de las columnas 'Cancion' + 'Artista'

'Duracion_min' tiene dos valores nulos por lo que decidimos eliminarlos. También observamos que tiene valores mayores que diez minutos y menores de 1 minuto. Decidimos eliminar las canciones con más de diez minutos porque en su mayoría eran audiolibros o 'nonstop remixes' y las canciones de menos de un minuto porque eran adelantos de canciones.

Aún existen valores nulos en las columnas "Visualizaciones_YT" y "Reproducciones_SP". Para abordar esta situación, decidimos calcular la relación entre las medias generales de "Visualizaciones_YT" y "Reproducciones_SP". Creamos un índice basado en esta relación y multiplicamos las "Visualizaciones_YT" de cada canción por este índice para estimar las "Reproducciones_SP" en aquellas canciones donde faltaba dicha información. De manera análoga, también realizamos el cálculo inverso.

LIMPIEZA DE DATOS POR VARIABLE

Hablabilidad:

Observamos que los valores con una "hablabilidad" muy alta corresponden a audiolibros, mientras que los valores con una "hablabilidad" muy baja corresponden a "White Noise" (ruido instrumental para relajación). Específicamente, aquellos con "hablabilidad" menor a 0.2 son por lo general "White Noise", y aquellos con "hablabilidad" mayor a 0.68 son audiolibros. Eliminamos estos valores porque no nos sirven.

Canción, Artista y Álbum.

Realizamos una búsqueda exhaustiva por nombres de estas columnas que nos parecían que podían contener datos que no fuesen buenos para el estudio y eliminamos las filas donde se encontraban estos valores. Por ejemplo, detectamos que algunas canciones y artistas incluyen términos como "White Noise" o "Meditation", que son más relevantes para fines de relajación y meditación que para nuestro análisis musical. Asimismo, identificamos álbumes con nombres como "Motion Picture", "Soundtrack", "Movie" o "HBO", que generalmente se refieren a bandas sonoras de películas o series de televisión.

Instrumentalidad:

Al realizar un resumen estadístico de la variable, observamos que los valores que se encuentran agrupados hasta el tercer cuartil son aproximadamente 0. Es importante tener en cuenta que la variable de instrumentalidad tiene un rango de valores entre 0 y 1. Sin embargo, al examinar algunos casos donde la instrumentalidad se supone que es 0, encontramos que estas canciones abarcan diversos géneros musicales.

Debido a la falta de calidad de los datos en esta variable, hemos tomado la decisión de eliminar la columna en el análisis posterior.

ANÁLISIS DE DATOS

ANÁLISIS UNIDIMENSIONAL

Análisis unidimensional de variables numéricas:

- Resumen estadístico para observar los principales datos estadísticos.
- Clasificación de las diez canciones con mayor y/o menor valor de la variable para ver cuales canciones representan más o menos a cada variable
- Histograma para visualizar la distribución y comportamiento de la variable
- Gráfico de caja para visualizar la mediana y los outliers
- Gráfico de violín para visualizar la densidad de la distribución de los valores
- Realizamos las pruebas Normal-test y Test Anderson-Darling para comprobar si la variable sigue una distribución normal. También realizamos el gráfico “qqplot” para verificarlo visualmente. Este gráfico compara los cuantiles de los datos observados con los cuantiles esperados de la distribución teórica.

Análisis unidimensional de variables no numéricas:

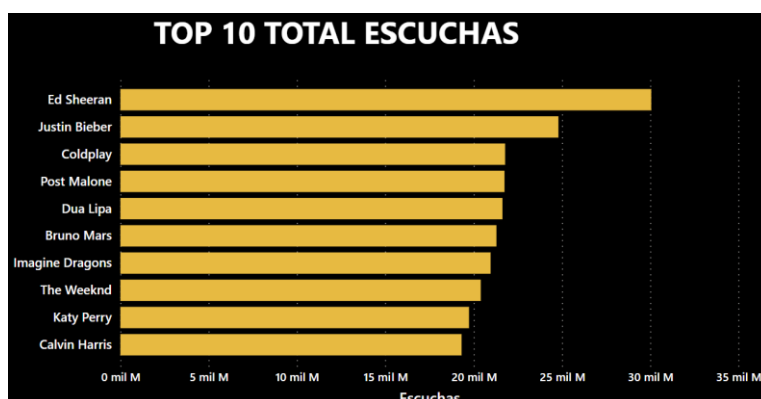
- Contamos los valores por cada tipo de álbum y lo graficamos en un gráfico de torta para visualizar que porcentaje representa cada tipo.

ANÁLISIS MULTIDIMENSIONAL

Top Artistas

Agrupamos los 10 artistas más escuchados de cada plataforma y en Escuchas_Totales y graficamos con un gráfico de barras horizontal para realizar una comparación entre los artistas más destacados.

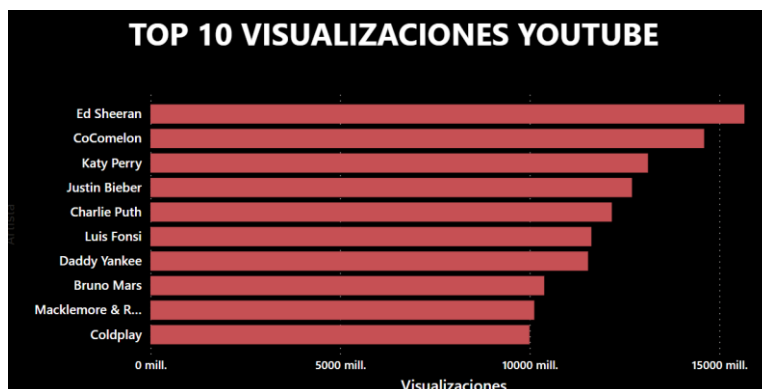
Top 10 total escuchas



2Top 10 Reproducciones Spotify



3Top 10 visualizaciones Youtube

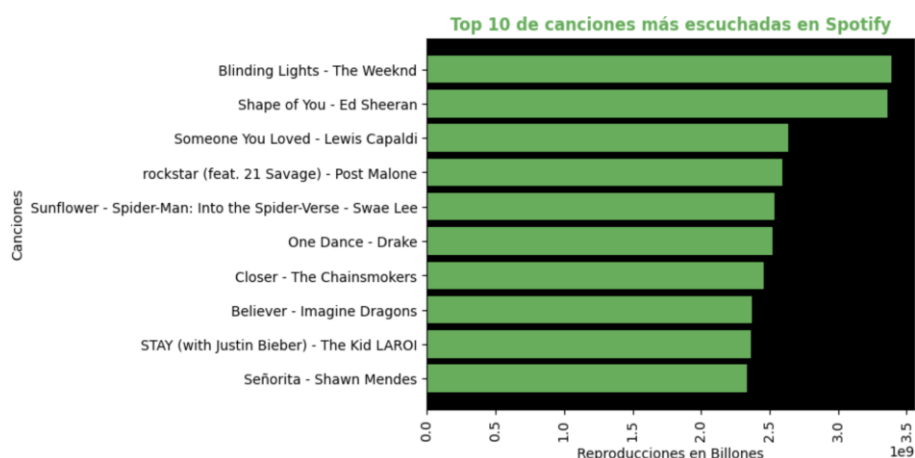


Observamos que algunos artistas se repiten en todas las listas, muchos van variando. El éxito en una plataforma no implica el mismo éxito en otra.

Top Canciones

Agrupamos las 10 canciones más escuchadas de cada plataforma y en la variable Escuchas_Totales y graficamos con un gráfico de barras horizontal para realizar una comparación entre las mejores canciones.

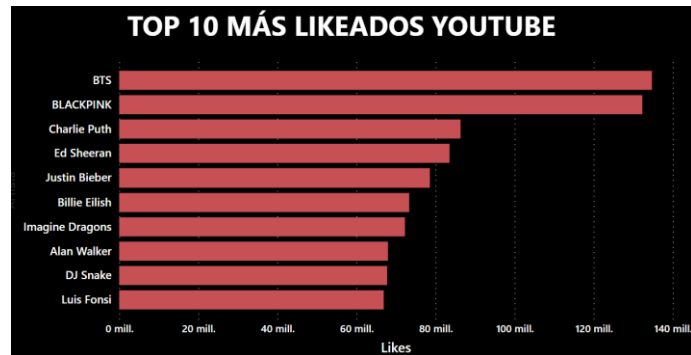
4Top 10 Escuchas Totales



Top likes

Agrupamos las 10 canciones más likeadas de cada Youtube y lo representamos en un gráfico de barras horizontal para realizar una comparación entre las más likeadas.

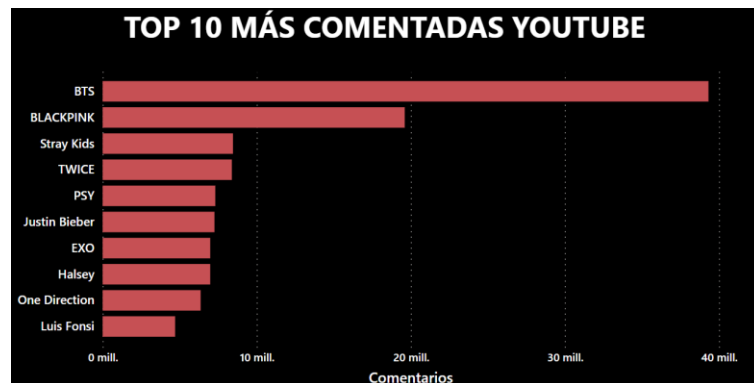
5Top 10 Likes



Top comentarios

Agrupamos las 10 canciones más comentadas de cada Youtube y lo representamos en un gráfico de barras horizontal para realizar una comparación entre las más comentadas.

6Top 10 comentarios Youtube

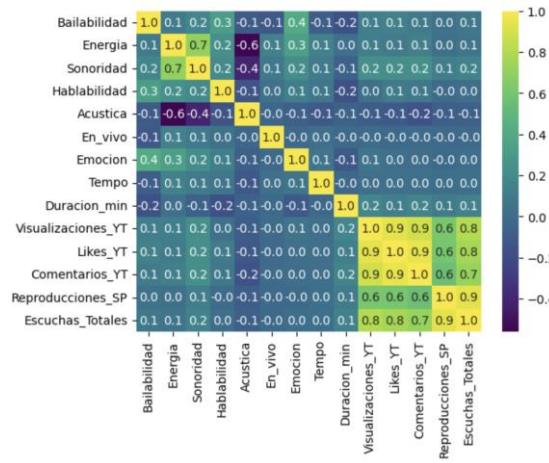


Hemos observado que los artistas asiáticos, especialmente los coreanos, generan un mayor compromiso o conexión con su público, a pesar de no ser los artistas más escuchados.

Análisis de correlación

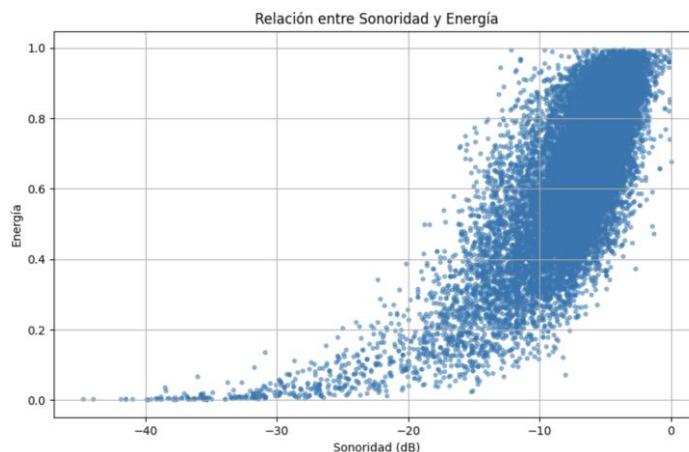
Realizamos la prueba de correlación de Spearman para evaluar la relación entre todas las variables numéricas. Optamos por la prueba de Spearman porque es una prueba no paramétrica y nuestras variables no siguen una distribución normal.

7Correlacion variables por canción



El gráfico nos muestra una clara relación entre las métricas de las canciones que son las Visualizaciones de Youtube, las Reproducciones de Spotify, los Likes y comentarios de Youtube y las escuchas totales. No se ven relaciones significativas entre las variables características de las canciones. Solo podemos observar una correlación entre la Sonoridad y la Energía de las canciones. Realizamos un gráfico de dispersión para ver cómo se comportan entre sí.

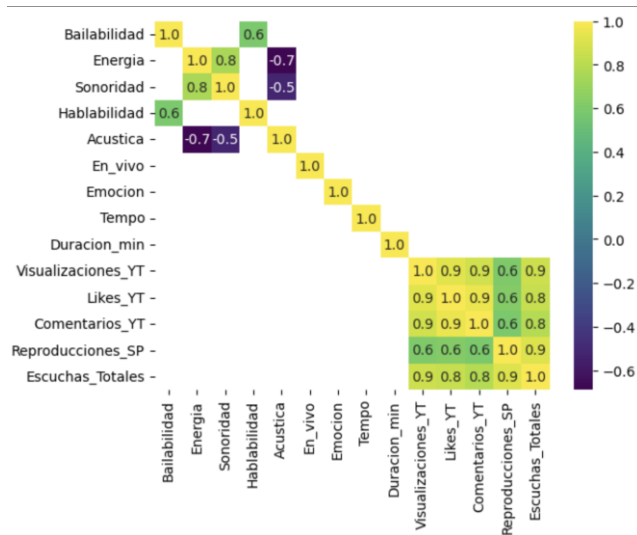
8Correlación Energía y Sonoridad



Observamos que cuanto más energética sea una canción, más ruidosa es. Los puntos se agrupan y siguen una dirección. Se puede ver la correlación.

Para profundizar en el análisis de correlación entre las variables numéricas de las canciones, llevamos a cabo una nueva estrategia. Decidimos agrupar las canciones por artista y calcular la media de cada variable para cada artista. De esta manera, pudimos examinar cómo se relacionan los artistas en función de las características promedio de sus canciones y métricas. Este enfoque nos proporciona una perspectiva más completa de las tendencias y similitudes entre los artistas en términos de las características musicales de sus canciones.

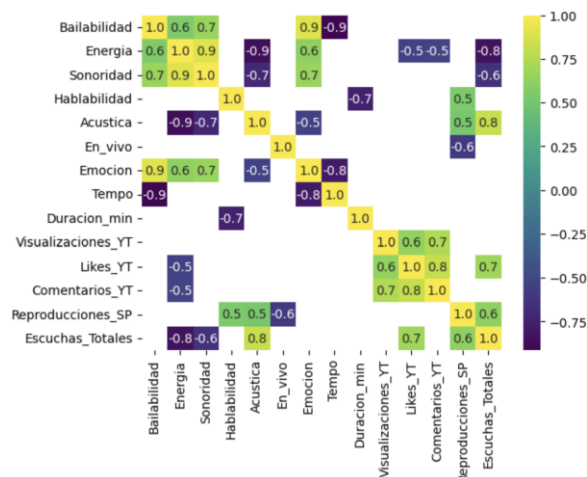
9Correlación variables por artista



En el gráfico actual, solo se muestran correlaciones superiores al 50%. Una observación interesante que se destaca en comparación con el análisis anterior es la relación inversa entre la Acústica y la Energía. Esto sugiere que cuanto más acústicas sean las canciones de un artista, menos energéticas serán.

Siguiendo con nuestro análisis, hemos decidido llevar a cabo una última prueba de correlación, pero esta vez enfocándonos exclusivamente en los 10 artistas más escuchados. Este enfoque nos permitirá examinar las relaciones específicas entre estas figuras destacadas en la industria musical y comprender cómo se relacionan en función de sus características musicales y sus métricas

10Correlación variable Top 10 artista



Podemos ver como la correlación aumenta entre los 10 artistas más escuchados. Se puede afirmar que entre ellos tienen características similares.

Hipótesis 1:

Las canciones más populares son las más bailables, alegres y enérgicas

Para contestar esta pregunta, analizaremos las medias de las variables bailabilidad, emoción y energía de la población y de una muestra de las 10 canciones más escuchadas.

Para el análisis, realizaremos una prueba de Hipótesis para saber si las medias poblacionales son iguales a las medias muestrales de las variables elegidas

Como hemos demostrado anteriormente, la Bailabilidad, la Emoción y la Energía no siguen una distribución normal, por lo que debemos realizar un test para variables no paramétricas

Prueba de Mann-Whitney

La prueba de **Mann-Whitney** la utilizamos para comparar dos grupos independientes y no podemos asumir que los datos siguen una distribución normal. Sirve para evaluar diferencias entre las medias. En este caso lo utilizaremos para comprobar si las medias de las tres variables mencionadas son iguales en la población y en la muestra.

En el caso del estudio de la Bailabilidad el resultado nos da un p_valor (0.1325904213177127) más alto que el nivel de significación (0.05) por lo que no podemos rechazar la Hipótesis nula. Las medias serían iguales.

Aunque la prueba nos dice que las medias no tienen una diferencia significativa, igualmente realizaremos un nuevo test para ver cuál es menor.

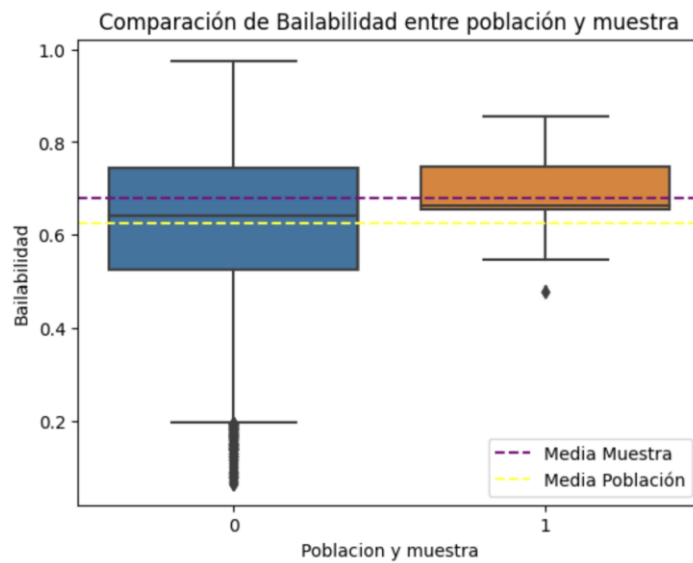
Planteamos como Hipotesis Nula que las canciones de la población son en media, igual o más bailables que las de la muestra del top 10

Y como Hipótesis Alternativa establece que las canciones de la población son en media, menos bailables que las de la muestra del top 10

El test nos dice nuevamente que no podemos rechazar la Hipótesis Nula. No hay suficiente evidencia para afirmar que las canciones de la población son en media, menos bailables que las de la muestra del top 10. $P_valor = 0.06629521065885635 > 0.05$.

Por último, para que quede claro de una manera más visual, realizamos un gráfico de caja para comparar las medias de bailabilidad.

11 Comparación de bailabilidad



De esta manera podemos ver una diferencia entre las medias de Bailabilidad. La media de la bailabilidad de la muestra es superior a la media de la Población

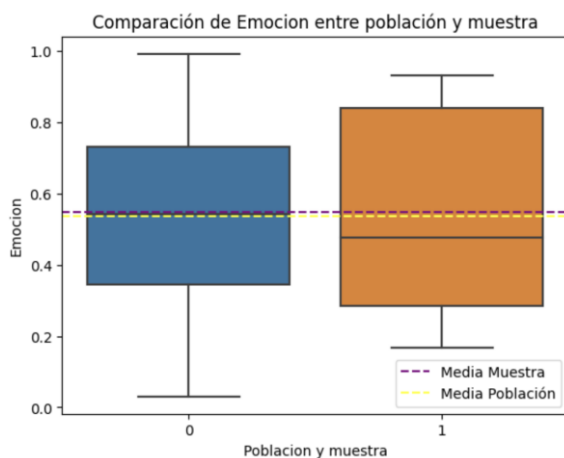
Conclusión: Las canciones del Top 10 son, en media, más bailables que las que no están en el Top 10. No tienen una diferencia significativa, pero suelen ser más bailables.

Realizamos estas mismas pruebas para las variables emoción y energía.

Se rechazaron las hipótesis nulas en todos los casos pero notamos diferencias en la comparación del gráfico de caja.

Emoción:

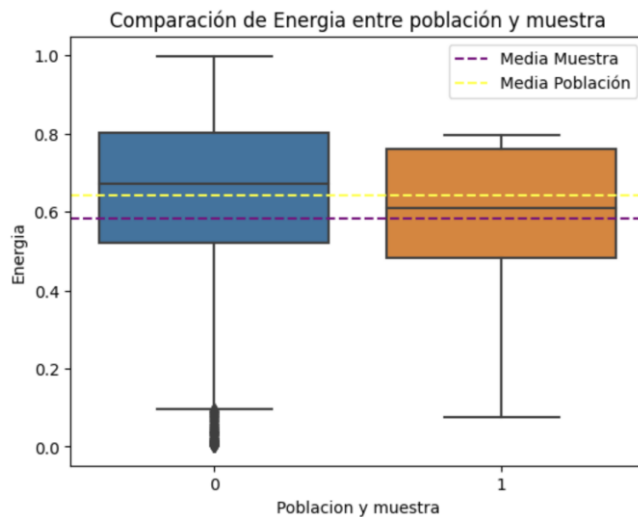
12 Comparación Emoción



De esta manera podemos ver como se nota una pequeña diferencia entre las medias de emoción. Las canciones del Top 10 no son necesariamente más alegres que las que no están en el Top 10

Energía:

13 Comparación energía



Con el gráfico, observamos diferencia entre las medias de Energía. La media de la Energía de la Población es superior a la media de la muestra

Conclusión Hipótesis 1:

Las canciones del Top 10 son, en media, menos enérgicas que las que no están en el Top 10. No tienen una diferencia significativa, pero suelen ser menos enérgicas

Hipótesis 2:

Las canciones más escuchadas no pueden tener una duración muy larga.

Para responder a esta pregunta, llevaremos a cabo un análisis de las variables "duración_minutos" y "Total_Escuchas". Para facilitar el estudio, categorizaremos la duración de las canciones en tres grupos:

Grupo 1: Canciones de 1 a 3 minutos.

Grupo 2: Canciones de más de 3 a 6 minutos.

Grupo 3: Canciones de más de 6 a 10 minutos.

Prueba de Kruskal-Wallis

Para el análisis de estas categorías, aplicaremos utilizaremos el test de **Kruskal-Wallis**, el cual es adecuado para comparar las medias de tres o más grupos dentro de una misma población.

En caso de obtener diferencias significativas entre las medias, procederemos a realizar el test de **Mann-Whitney** para determinar qué grupo muestra un mayor número de escuchas.

Hipótesis Nula: las medias de los 3 grupos son iguales

Hipótesis Alternativa: Las medias de los 3 grupos son distintas

El test de Kruskal-Wallis nos arroja el siguiente resultado:

Rechazamos la hipótesis Nula. Hay diferencias significativas entre los grupos. $P_valor = 1.7380793728894112e-122 < 0.05$

Prueba de Mann-Withney

Realizamos el test de MannWithney entre los distintos grupos para evaluar cada grupo entre sí y llegamos a los siguientes resultados:

los grupos 1 y 2 son diferentes

El grupo 1 es superior a 2

los grupos 1 y 3 son diferentes

El grupo 1 es superior a 3

los grupos 2 y 3 son diferentes

El grupo 2 es inferior a 3

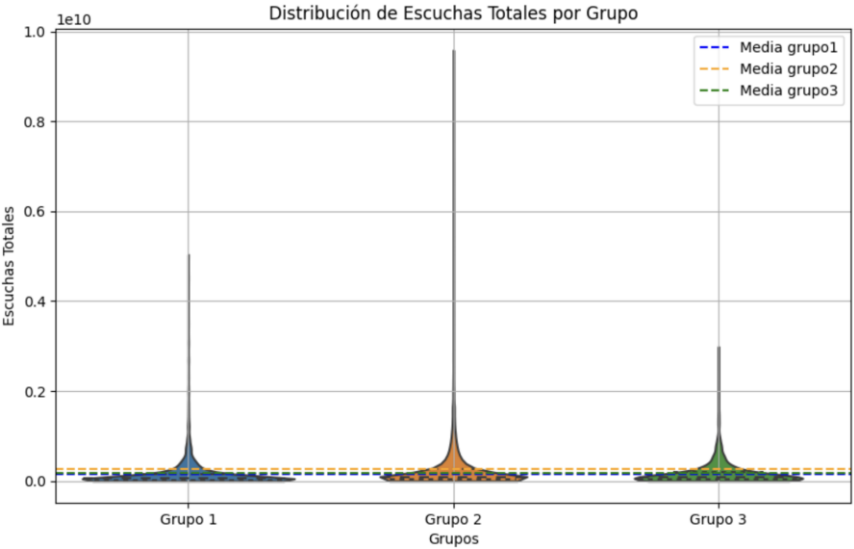
Conclusión Hipótesis 2:

Las canciones con la duración entre 1 y 3 min son las más escuchadas.

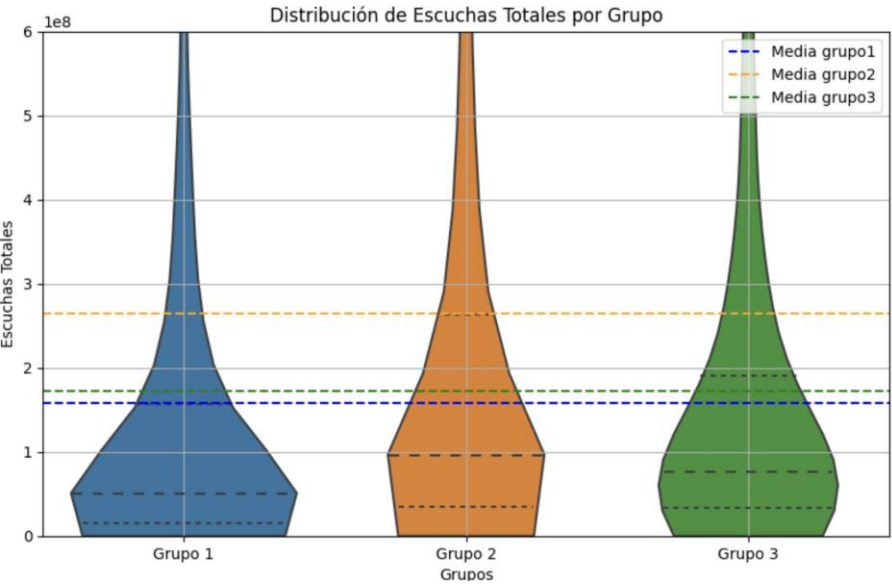
Sin embargo, las canciones con más de 6 minutos tienen más escuchas que las que tienen entre 3 y 6 minutos.

Gráfico para ver las diferencias:

14 Distribución de Escuchas Totales por Grupo



15 Distribución de Escuchas Totales por Grupo aumentada



REFERENCIAS

Guía para ser un Data Scientist – Daniel Montes Serrano - 2023

"Música". En: Significados.com. Disponible en: <https://www.significados.com/musica/> - 22/01/2021.

CONCLUSIONES

Los resultados obtenidos en este estudio proporcionan información valiosa para comprender la dinámica de la industria musical digital. Se ha confirmado que las canciones más populares suelen ser más bailables y alegres, aunque no necesariamente más enérgicas que la media de las canciones, aunque estas diferencias no resultan significativas. Además, se ha destacado la preferencia del público por canciones de duración corta, con una duración promedio de uno a tres minutos.

También hemos llegado a la conclusión de que la popularidad en una plataforma está relacionada con la popularidad en otra, aunque no garantiza que un artista o canción sea igualmente popular en ambas. Observamos que los artistas coreanos generan una alta interacción y conexión con el público, aunque eso no los ponga en las listas de los más escuchados.

Estos hallazgos nos brindan una visión detallada de cómo el público interactúa con la música en las plataformas de streaming y cómo ciertas características musicales influyen en la preferencia de las canciones.

ANEXOS

REFERENCIAS DE COLUMNAS

Canción	Nombre de la canción, tal y como aparece en la plataforma Spotify.
Artista	Nombre del artista.
Url_spotify	Url del artista en Spotify
Álbum	El álbum en el que se encuentra la canción en Spotify.
Tipo_Album	Indica si la canción se publica en Spotify como single o está incluida en un álbum.
Uri	El Url de la canción en spotify.
Bailabilidad	Describe lo adecuada que es una canción para bailar basándose en una combinación de elementos musicales como el tempo, la estabilidad del ritmo, la fuerza del compás y la regularidad general. Un valor de 0,0 es el menosailable y 1,0 el másailable.
Energía	Va de 0,0 a 1,0 y representa una medida perceptiva de intensidad y actividad. Normalmente, las pistas energéticas son rápidasy ruidosas. Por ejemplo, el death metal tiene mucha energía, mientras que un preludio de Bach puntúa bajo en la escala. Las características perceptivas que contribuyen a este atributo incluyen el rango dinámico, el volumen percibido, el timbre, la velocidad de inicio y la entropía general.
Key	La clave en la que se encuentra la pista. Los números enteros se asignan a los tonos utilizando la notación estándar de Clase de Tono. Por ejemplo, 0 = Do, 1 = Do#/Reb, 2 = Re, y así sucesivamente. Si no se detectó ninguna clave, el valor es -1.
Sonoridad	Se refiere al volumen o nivel de intensidad global de una pista musical y se mide en decibelios (dB). Los valores de sonoridad se promedian en toda la pista para obtener una medida representativa del volumen general. Esta medida es útil para comparar el volumen relativo entre diferentes pistas. La sonoridad es la cualidad de un sonido que se asocia psicológicamente con su intensidad física, que se representa mediante la amplitud de la onda sonora. En otras palabras, la sonoridad

	<p>es la percepción subjetiva de la fuerza o el nivel de volumen de un sonido. Los valores de sonoridad típicamente oscilan entre -60 y 0 dB, donde los valores más cercanos a 0 dB indican un volumen más alto, y los valores más cercanos a -60 dB indican un volumen más bajo.</p>
Hablabilidad	<p>Detecta la presencia de palabras habladas en una pista. Cuanto más exclusivamente se asemeje la grabación a palabras habladas (por ejemplo, un programa de entrevistas, un audiolibro, poesía), más cercano a 1.0 será el valor del atributo. Los valores superiores a 0.66 describen pistas que probablemente están compuestas completamente por palabras habladas. Los valores entre 0.33 y 0.66 describen pistas que pueden contener tanto música como palabras habladas, ya sea en secciones o superpuestas, incluyendo casos como la música rap. Los valores inferiores a 0.33 representan probablemente música y otras pistas que no son similares a palabras habladas.</p>
Acústica	<p>Sirve para medir la probabilidad de que una pista sea acústica. Es una medida de confianza que varía de 0.0 a 1.0. Un valor de 1.0 indica una alta confianza de que la pista es completamente acústica, es decir, que se interpreta utilizando instrumentos musicales acústicos en lugar de instrumentos electrónicos o amplificadores.</p>
Instrumentalidad	<p>Se utiliza para estimar si una pista es instrumental o contiene voces. Valores cercanos a 1.0 indican una alta probabilidad de que la pista sea completamente instrumental, mientras que valores por encima de 0.5 también sugieren pistas instrumentales, pero con mayor confianza a medida que se acercan a 1.0.</p>
En_vivo	<p>Detecta la presencia de una audiencia en la grabación. Valores más altos de liveness representan una mayor probabilidad de que la pista haya sido interpretada en vivo. Un valor por encima de 0.8 indica una fuerte probabilidad de que la pista sea en vivo.</p>
Emoción	<p>es una medida de 0.0 a 1.0 que describe la positividad musical transmitida por una pista. Las pistas con un alto valor de valence suenan más positivas (por ejemplo, alegres, felices, eufóricas), mientras que las pistas con un bajo valor de valence suenan más negativas (por ejemplo, tristes, deprimidas, enojadas).</p>

Tempo	Es el tempo estimado en general de una pista, medido en pulsaciones por minuto (BPM beats per minute). En terminología musical, el tempo se refiere a la velocidad o ritmo de una pieza musical y se deriva directamente de la duración promedio de los pulsos. Se calcula a partir de la duración promedio de los pulsos o golpes que conforman el ritmo de una pista. Por ejemplo, una pista con un tempo alto tendrá un ritmo rápido, mientras que una pista con un tempo bajo tendrá un ritmo más lento.
Duracion_min	Duración de la pista en minutos.
Reproducciones_SP	Número de streams de la canción en Spotify.
Url_youtube	Url de la canción en Youtube.
Titulo_video	Título del videoclip en Youtube.
Canal	Nombre del canal que ha publicado el vídeo.
Visualizaciones_YT	Número de vistas.
Likes_YT	Número de likes.
Comentarios_YT	Número de comentarios.
Descpcion	Descripción del video de Youtube.
Licencia	Indica si el video representa contenido con licencia, lo que significa que el contenido fue subido a un canal vinculado a un socio de contenido de YouTube y luego reclamado por ese socio.
Oficial_video	Valor booleano que indica si el video encontrado es el video oficial de la canción.