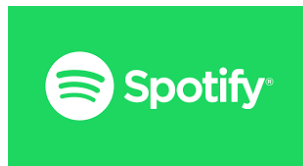


Proyecto final



CURSO DATA ANALYTICS

Franco Mauricio Blanco

CODERHOUSE

Año 2025

Descripción del tema

En el presente trabajo se analizará el éxito y la cantidad de reproducciones de canciones en Spotify y Youtube interpretadas por artistas de primer nivel. Se buscará establecer relaciones entre estos indicadores y diversas características organizativas y musicales, con el fin de identificar patrones comunes en las canciones más reproducidas y los videos más reproducidos. El dataset seleccionado contiene información de más de 20 mil de canciones obtenidas a partir de la API de Spotify y youtube. A pesar de que los datos fueron actualizados por última vez hace 2 años, continúan siendo relevantes para comprender las tendencias del mercado musical. Este conjunto de datos abarca una variedad de géneros y proporciona detalles clave sobre cada canción, excepto la fecha de salida de los videos y las canciones. El objetivo principal de este estudio es proporcionar una comprensión más profunda de los factores que influyen en el número de reproducciones de una canción tanto en audio como en video. Con ello, se busca ofrecer información útil para orientar proyectos musicales y ayudar a los artistas en la toma de decisiones estratégicas a lo largo de su carrera.

Índice

1. Introducción	3
1.1. Objetivo	3
1.1.1. Nivel de análisis	3
1.2. Base de datos	3
1.3. Hipótesis	5

Capítulo 1

Introducción

1.1. Objetivo

El objetivo principal de este trabajo es analizar la correlación entre el número de reproducciones de una canción y diversas características musicales y organizativas. A partir de este análisis, se pretende generar recomendaciones basadas en datos para orientar a los artistas y productores en el diseño de estrategias que optimicen la difusión de sus canciones y maximicen su alcance en ambas plataformas (youtube y spotify).

1.1.1. Nivel de análisis

Este estudio aspira a un nivel de análisis táctico-estratégico ya que busca identificar tendencias en la relación entre características musicales y reproducciones para el beneficio de pequeños y grandes artistas junto con empresas de streaming.

1.2. Base de datos

La base de datos consta de un archivo csv titulado **Spotify_Youtube.csv**. Se eligió este dataset por sobre los demás debido al gran número de indicadores junto con el número de reproducciones tanto en youtube como en spotify. Ambas métricas permiten relacionar las características de las canciones con el número de reproducciones. Si bien los otros dos datasets superan ampliamente el número de canciones

este es el único que encontré que reunía todas las características deseadas.

El mismo cuenta con información para alrededor de 20 mil canciones. Para cada una de las canciones el dataset brinda:

- Nombre de la canción.
- Nombre del artista que tocó la canción.
- Url spotify la dirección de la canción en spotify
- Nombre del álbum al que pertenece la canción.
- Albumtype Indica si la canción se lanzó como un single o un album.
- Uri link para encontrar la canción dentro del spotify utilizando el API.
- Número de canción (dentro del álbum).
- Número del disco dentro del historial de la banda.
- Danceability mide a la canción en cuanto a su bailabilidad. Entrega una probabilidad de que tan bailable es la canción basado en varias características musicales.
- Energy es una medida que tan movida es la canción. Esto depende del tempo, rango dinámico y otros factores. También es un número entre 0 y 1.
- Key le asigna una tonalidad a la canción según un numero entero del 0 al 11. Si no detecta la tonalidad asigna -1.
- Loudness mide el valor promedio de la canción en decibeles.
- Mode indica si la canción está en modo mayor o menor, toma solo dos valores: 0 o 1.
- Speechiness mida la proporción de diálogo respecto a música que existe en la canción. Este número es también del 0 al 1 donde 1 representa una pista compuesta enteramente por palabras mientras que un valor de 0 representa una pista instrumental.
- Acousticness Mide la probabilidad de que una pista sea instrumental. Cuanto más cerca esté el valor de 1, más probable es que la pista no tenga contenido vocal.

- Liveness Mide si la pista fue grabada en vivo, con la presencia de una audiencia. Los valores más altos indican una mayor probabilidad de que la pista sea en vivo.
- Valence Mide la "positividad" de una pista, desde 0 hasta 1. Las canciones con alta valencia suenan más alegres y felices, mientras que las de baja valencia suenan más tristes o melancólicas.
- Tempo Indica la velocidad o ritmo de una pista, medida en pulsos por minuto (BPM). Un valor más alto significa que la pista es más rápida, mientras que un valor más bajo indica un ritmo más lento.
- Duration (ms) La duración de la canción en milisegundos
- Stream el número de reproducción de dicha canción en spotify.
- Urlyoutube El link de la canción en youtube.
- Title título del videoclip en youtube.
- Channel nombre del canal en el que se subió el video.
- Views número de visitas de dicho videoclip.
- Likes número de likes del video de youtube.
- Comments número de comentarios.
- Description descripción del video de youtube.
- Licensed indica si el video está bajo licencia de un canal de youtube o discográfica.
- officialvideo valor de 0 o 1 para indicar si el video corresponde al video oficial.

1.3. Hipótesis

El presente trabajo postula la existencia de un impacto significativo de ciertas características musicales en el número de reproducciones para grandes artistas. En particular: danceability, energy, mode, loudness, speechiness, valence y tempo. Para ser preciso se espera que el impacto de cada uno de los factores sea el siguiente:

danceability: A mayor danceability, mayor será la cantidad de reproducciones.

energy: de la misma manera se espera que una canción que cuenta con mayor energía tenga asociado un mayor número de reproducciones.

mode: Se espera que haya un sesgo hacía el modo mayor (alegre) inclinado hacia el modo mayor alegre.

loudness: se espera que una canción con mayor volumen sea más escuchada.

speechiness: se espera que haya un número óptimo de speechiness tal que favorezca el número de escuchas.

valence: Al igual que el modo, se espera que un valor de valencia más cercano al 1 tenga resultados positivos en cuanto a reproducciones.

Por último, se espera que haya un rango de tiempo óptimos para los cuales el número de canciones tengan un buen número de reproducciones.

Cómo se ve arriba se espera que varios de estos factores tengan un impacto positivo en el número de streams y reproducciones. Para estudiar este impacto espero obtener correlaciones positivas entre el número de reproducciones y estas características.