EDA Project

Daniel Carrera Cabezuelo

DATA SCIENCE – THE BRIDGE

**¿Quién es el artista, o grupo musical, más importante de nuestra historia reciente?**

Un análisis comparativo de datos.

# Abstract

La música. El cuarte arte clásico de la humanidad. Desde tiempos inmemoriales, la música ha sido uno de los pasatiempos más importantes del ser humano. Desde las civilizaciones clásicas, o incluso antes, pasando por los juglares de la edad media, alcanzó su apógeo a inicios del siglo XVI, con los grandes músicos del barroco y del renacimiento, la conocida música clásica. ¿O no? Tal vez los más puristas coincidirán en que es así, otros dirán que el apogeo llegaría en el siglo XX, y si sigues preguntando, cada persona defendera una década distinta. Y si preguntas a las generaciones más nuevas, puede que cualquier cosa que no haya salido en este siglo les parezca arcaico o directamente lo desconozcan. Cual es la época dorada de la música es un eterno debate. Y si buscamos al mejor artista musical, el debate es aún mayor. La música va en gustos personales, desde luego. Lo que esta claro, es que a nivel subjetivo, todos tenemos preferencia por un estilo o una época musical. Pero vivmos en la gran era de los datos, ¿no? ¿Por que no vemos si estos pueden arrojar un poco de luz, de la forma más objetiva posible, a esta pregunta?

# Planteamiento de la hipótesis

El progreso del hombre siempre ha ido marcado por hacerse ua pregunta y buscar una respuesta. Vamos a plantearnos la hipótesis principal:

*¿Pueden los datos arrojarnos luz sobre el debate de quien es el artista o grupo musical más importante de todos los tiempos?*

Esta es la gran cuestión. Pero es una cuestión demasiado compleja para poder buscar una respuesta directa. Asique intentemos establecer una serie de premisas, o puntos de partida, sobre los que poder cimentar nuestra búsqueda.

* **VENTAS**: Este es uno de los parámetros más descriptivos. Las ventas que consigue un músico con sus álbumes o sus canciones son la forma más objetiva de ver la acogida que tiene dicho artista entre el público. Será interesante analizar las tendencias de las ventas a lo largo de los años y ver como se comportan con la aparición de nuevas tecnologías.
* **GALARDONES**: Las instituciones más importantes de asociaciones de música en el mundo llevan décadas premiando con diferentes premios a los mejores artistas, canciones, o álbumes, ya sea en base a ventas, votaciones o apoyo de la gran masa. Será interesante analizar a los artistas más galardonados en las diferentes premiaciones a lo largo de los años.
* **LISTAS DE ÉXITOS**: Todos los músicos anhelan alcanzar la cima de las listas de éxitos y tener la mayor visibilización y proyección posibles, pocos lo alcanzan, pero solo los mejores lográn llegar y mantenerse en lo más alto. Será interesante analizar quienes han conseguido mantenerse en este selecto grupo a lo largo del tiempo, ver cuanto tiempo han estado en la cima y con cuantas canciones lo han logrado.

# Antes de empezar

Una vez establecidas las directrices de lo que queremos encontrar, llega el momento de ponerse manos a la obra y salir en busca de los datos necesarios para poder realizar nuestra investigación.

Antes de comenzar, debemos dejar algunas temas claros:

* Esto es un analísis exploratorio de datos de un futuro data scientist, que esta empezando, donde se intentará dar una respuesta lo más veraz y objetiva posible, pero no es una verdad absoluta. Estoy abierto a cualquier crítica constructiva y capacidad de mejora.
* Además, existe el hecho de que, por desgracia, hay épocas de las que carecemos de datos, (no sabemos cuanta gente escuchaba a Mozart o cuantos discos podría haber vendido Bethoveen). Este sesgo insalvable es algo por lo que el hecho de declarar al músico más importante de todos los tiempos es, en esencia, algo capcioso, puesto que solo podremos analizar datos de la música del siglo XX y XXI.

# Búsqueda de los datos:

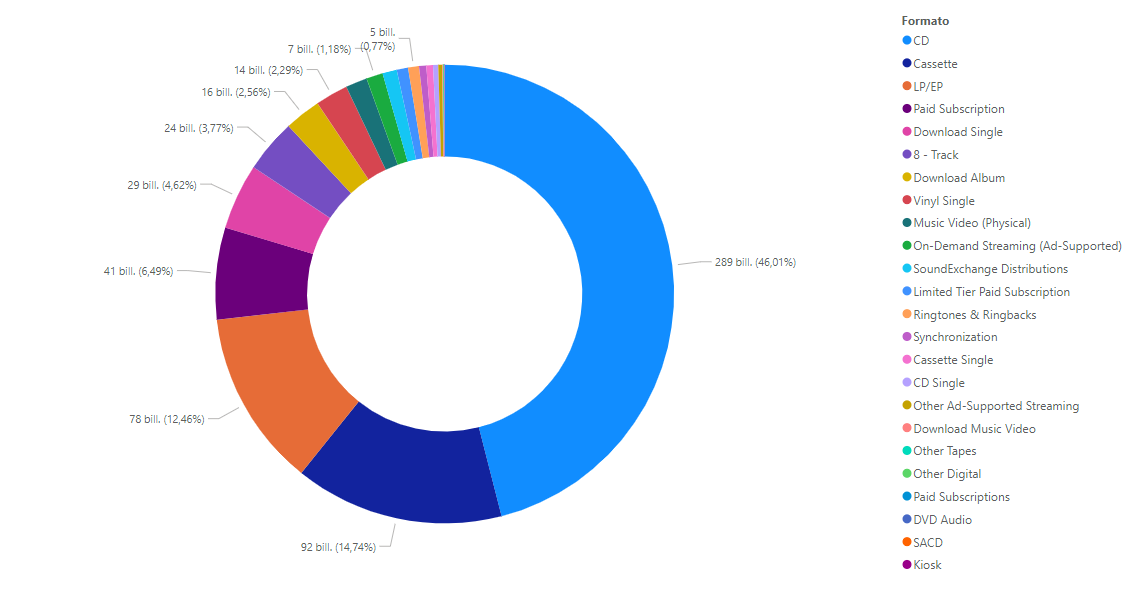
Vamos a seguir los tres pilares que hemos designado a la hora de buscar los datos. Empezaremos por las ventas. Lo primero que debemos tener en cuenta, es que los tiempos cambian, y la música también. Así que ver como ha variado el mercado músical a lo largo de los años es buena opción. Para ello, podemos utilizar este database de Kaggle, <https://www.kaggle.com/code/bourrinemusique/music-sales-riaa/input> . Este database muestra los diferentes formatos en los que se ha producido y vendido la música desde la década de los 70 hasta el año 2019. Vemos que hay un sesgo importante, pues no tenemos datos de ventas hasta el año 1973, pero los datos que nos proporciona la RIAA (Recording Industry Association of America), de la cual se han extraido, son a partir de dicha década. Haciendo un poco de limpieza y analizando los datos podemos ver:

Gráfico, Gráfico de barras, Histograma

Descripción generada automáticamente

Por lo que podemos ver, ha habido un cambio de tendencia en la forma de como se crea la música, como se distribuye y en definitiva, como se vende. Podemos ver que durante la década de los 70 y 80, la forma de consumir música era, puramente, a través del formato físico: Cassettes, 8-Tracks (Cartuchos de 8 pistas), LP/EPs y Vinilos de una sola canción. En 1982, con la creación del CD (Compact Disc) se inicia un cambio de tendencia que a partir de los 90 practicamente monopolizaría la industria músical y se convertiría hasta hoy, como veremos en el siguiente gráfico, en la mayor forma de venta y recaudación de la historia de la música.

Este monopolío del CD comenzaría su declive en 1999 con la aparición de Napster, la primera gran red P2P de intercambio donde la gente podía compartir sus colecciones de música en formato digital (MP3). La aparición de plataformas como iTunes y YouTube favorecieron a la digitalización del formato músical que alcanzaría el punto de inflexión con Spotify en 2008, donde el formato digital y en concreto, el formato de suscripción de pago para la reproducción de música cambió el paradigma que había hasta entonces. La industria se ha ido adaptando desde entonces con al aumento de ‘Singles’ en solitario en estas plataformas y la disminución del formato “album”.



A pesar de ello, y como podemos comprobar en esta gráfica que nos muestra la recaudación historica de la industria por formatos, vemos que el formato físico a día de hoy sigue siendo el culpable del más del 75% de la recaudación de los músicos, siendo el CD el mayor valor de todos.

Una vez creado este contexto inicial, ¿qué información nos da todo esto con respecto a nuestra pregunta objetivo? Pues bastante información útil para entender, por ejemplo, que a la hora de ver las ventas por artistas, que es el siguiente punto que valoraremos, es muy probable que veamos un mayor número de artistas de las últimas decadas del siglo XX y de la generación del 2000 debido a que en muchos casos toda su música se distribuia en formato físico, con el encarecimiento que ello conlleva con respecto al formato digital.

En esta siguiente gráfica vemos al TOP 25 de toda la historia de la música en cuanto a recaudación en ventas se refiere. Estos datos han sido extraidos mediante web scraping desde la página web: <https://chartmasters.org/best-selling-artists-of-all-time/> . Es una database muy actualizada, de Marzo de 2023.

Gráfico, Gráfico de barras

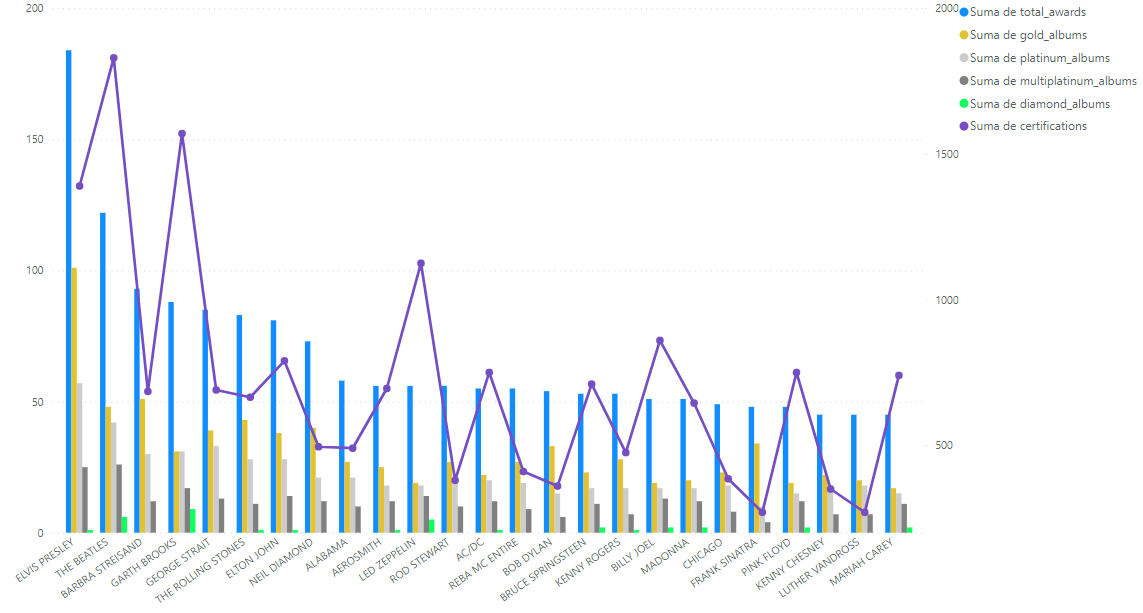
Descripción generada automáticamente

Esta gráfica, por temas de visualización y para lo que nos interesa, contando con ventas físicas y digitales de diferentes formatos nos da a los artistas con más ventas. Nos quedaremos con los cinco primeros de este apartado para el veredicto final.

|  |  |
| --- | --- |
| 1 | The Beatles |
| 2 | Michael Jackson |
| 3 | Elvis Presley |
| 4 | Queen |
| 5 | Madonna |

Continuamos con el siguiente punto a tener en cuenta, los galardones. Existen numerosos premios concedidos por diferentes organismos especializados a los músicos, ya sea por sus álbumes o por sus singles. Nos centraremos en los más prestigiosos, los galardones a un disco, que pueden ser de Oro, Platino, Multi Platino y Diamante, y los premios Grammy.

Empezaremos con los galardones a un disco. Nos apoyaremos en los concedidos por la RIAA (Recording Industry Association of America), el organismo principal que adjudica estos premios. Haciendo web scraping a su página: <https://www.riaa.com/gold-platinum/?tab_active=awards_by_artist#search_section> , podemos encontrar a los 150 artistas con más certificaciones de ventas, y cuantos galardones de este tipo han obtenido.



Vemos que no hay correlación directa entre el número de galardones y las ventas, por eso es importante analizarlo por separado. En este caso vemos que Elvis Presley es el que más discos galardonados tiene con diferencia y aun así no tiene más ventas que The Beatles.

Ahora vamos con los premios Grammy. Nos interesa saber cuales son los artistas con mayor número de Grammys a nivel histórico. Pero hay dos variantes, los ‘estándar’ de siempre y los latinos. Para lo que nos ocupa, los trataremos por igual. Haciendo web scraping a las paginas de wikipedia: <https://es.wikipedia.org/wiki/Anexo:R%C3%A9cords_de_los_Premios_Grammy> y <https://es.wikipedia.org/wiki/Anexo:R%C3%A9cords_de_los_Premios_Grammy_Latinos> hemos podido obtener lo siguiente:

Gráfico, Gráfico de barras

Descripción generada automáticamente

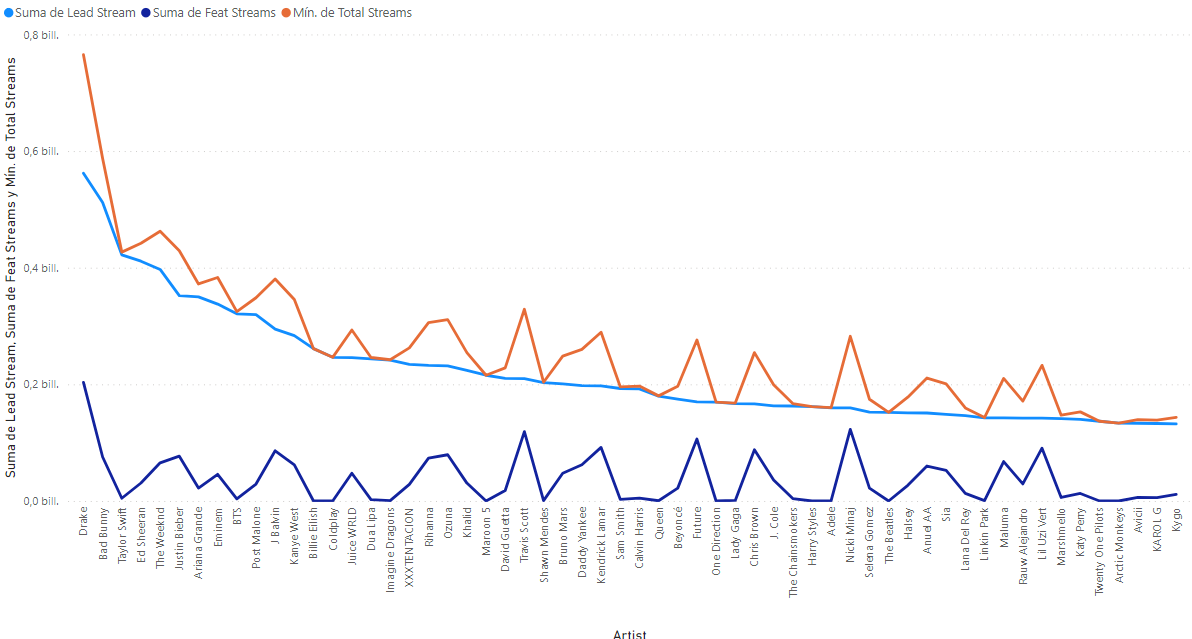
En resumen, el top 5 de Gold-Platinum Awards es:

|  |  |
| --- | --- |
| 1 | Elvis Presley |
| 2 | The Beatles |
| 3 | Barbara Streisand |
| 4 | Garth Brooks |
| 5 | George Strait |

Y el top 5 de los Grammys Históricos es:

|  |  |
| --- | --- |
| 1 | Beyoncé |
| 2 | Georg Solti |
| 3 | Quincy Jones |
| 4 | Alison Krauss |
| 5 | Juan Luis Guerra |

Y ya solo nos queda el último apartado, las listas de éxitos. A través de estes dataset de Kaggle: <https://www.kaggle.com/datasets/marshuu/best-artists-dataset-grammys-billboard-rs> , hemos podido obtener las lista de los mejores artistas según la revista Rolling Stones, de este otro de aquí: <https://www.kaggle.com/datasets/dhruvildave/billboard-the-hot-100-songs> , el historico de canciones de The Hot 100 de la revista Billboard y haciendo web scraping a esta página: <https://chartmasters.org/most-streamed-artists-ever-on-spotify/> , la lista de los 100 cantantes más escuchados en la historia de Spotify. Analizando y limpiando los datos podemos ver que:



En esta gráfica podemos ver a los artistas más escuchados en Spotify desde el nacimiento de esta plataforma. Podemos ver que el mayor peso de las reproducciones de cada artista es casi por completo de las canciones en solitario del propio cantante, pero podemos ver una tendencia de que los artistas que más reproducciones tienen tambien tienden a ser los que más canciones colaborativas de éxito tienen. Podemos ver que en estos casos los artistas con más reproducciones, evidentemente, son de las nuevas generaciones, en especial de comienzos de este siglo, a parte de porque las plataformas son muy actuales, el siglo pasado no existian, los cantantes nuevos tienden a sacar mayor cantidad de canciones que antes, tanto en solitario como en colaboración, cosa que en el siglo pasado no era tan frecuente.

Al final podemos ver que la clasificación según la revista Rolling Stones es:

|  |  |
| --- | --- |
| 1 | The Beatles |
| 2 | Bob Dylan |
| 3 | Elvis Presley |
| 4 | The Rolling Stones |
| 5 | Chuck Berry |

La de Spotify es:

|  |  |
| --- | --- |
| 1 | Drake |
| 2 | Bad Bunny |
| 3 | Taylor Swift |
| 4 | Ed Sheeran |
| 5 | The Weeknd |

Y la de la revista Billboard es:

|  |  |
| --- | --- |
| 1 | Taylor Swift |
| 2 | Elton John |
| 3 | Madonna |
| 4 | Drake |
| 5 | Kenny Chesney |

# Veredicto Final

Bueno, pues con todos estos datos, y gracias a nuestro querido amigo Python, hemos hecho un largo procesado de estos y hemos creado una forma de puntuación basada en dar la puntuación más alta al indice más bajo de cada dataframe y sumarlos todos para ver quien es el artista más influyente e importante de todos los tiempos (podrás ver el código integro de todo el procesado en mi Github). Y el artista, o grupo de música, más influyente de todos los tiempos es:

**THE BEATLES**

**THE ROLLING STONES**

**ELVIS PRESLEY**

# Conclusión

Pues, The Beatles, la banda de rock británico de la década de los 60 es la más influyente de todos los tiempos. Como ya hemos dicho, esto es lo que arrojan los datos, pero los gustos son subjetivos, y cada persona tendrá a su cantante favorito en base a sus vivencias con la música y lo que ésta te haga sentir.