

# Streaming - Hipóteses e Insights

**Cliente:** Gravadora

**Analistas de Dados:** Cristiane Thiel

Links: [Apresentação](#) - [Dashboard](#) - [Repositório no GitHub](#)

## Objetivo

O objetivo deste projeto é validar hipóteses sobre os fatores que influenciam o sucesso de uma música nas plataformas de streaming para apoiar uma gravadora no lançamento estratégico de um novo artista no mercado musical.

**Ferramentas e Tecnologias:** SQL, BigQuery, Power BI, Power Query, DAX, Python e VS Code.

## Problema Central

Como a gravadora pode usar dados de desempenho musical em plataformas como Spotify, Deezer e Apple Music para identificar os principais fatores que contribuem para o sucesso de uma música e, assim, orientar o lançamento de um novo artista?

A gravadora levantou uma série de hipóteses sobre o que faz uma música ser mais ouvida. Essas hipóteses incluem:

1. Músicas com BPM (Batidas Por Minuto) mais altos fazem mais sucesso em termos de número de streams no Spotify.
2. As músicas mais populares no ranking do Spotify também possuem um comportamento semelhante em outras plataformas, como a Deezer.
3. A presença de uma música em um maior número de playlists está correlacionada com um maior número de streams.
4. Artistas com um maior número de músicas no Spotify têm mais streams.
5. As características da música influenciam o sucesso em termos de número de streams no Spotify.

## Possíveis Stakeholders

- **Marketing:** busca entender quais características tornam uma música mais atrativa para otimizar campanhas de divulgação e aumentar o alcance.
- **Diretoria Executiva:** espera decisões estratégicas orientadas por dados para maximizar o investimento no lançamento do novo artista.

# Processamento e Análise

Foram disponibilizados três arquivos em formato csv. **Tabelas:** track\_in\_spotify, track\_in\_competition e track\_technical\_info.

Campo	Descrição
track_id	Identificador exclusivo da música.
track_name	Nome da música.
artist_name	Nome do(s) artista(s) da música.
artist_count	Número de artistas que contribuíram na música.
released_year	Ano em que a música foi lançada.
released_month	Mês em que a música foi lançada.
released_day	Dia do mês em que a música foi lançada.
in_spotify_playlists	Número de playlists do Spotify em que a música está incluída.
in_spotify_charts	Presença e posição da música nas paradas do Spotify.
streams	Número total de streams no Spotify. Representa o número de vezes que a música foi ouvida.

Campo	Descrição
track_id	Identificador exclusivo da música.
in_apple_playlists	Número de playlists da Apple Music em que a música está incluída.
in_apple_charts	Presença e classificação da música nas paradas da Apple Music.
in_deezer_playlists	Número de playlists do Deezer em que a música está incluída.
in_deezer_charts	Presença e posição da música nas paradas do Deezer.
in_shazam_charts	Presença e classificação da música nas paradas da Shazam.

Campo	Descrição
track_id	Identificador exclusivo da música.
bpm	Batidas por minuto, uma medida do tempo da música.
key	Tom musical da música.
mode	Modo de música (maior ou menor).
danceability_pct	Porcentagem que indica o quão apropriada a canção é para dançar.
valence_pct	Positividade do conteúdo musical da música.
energy_pct	Nível de energia percebido da música.
acousticness_pct	Quantidade de som acústico na música.
instrumentalness_pct	Quantidade de conteúdo instrumental na música.
liveness_pct	Presença de elementos de performance ao vivo.
speechiness_pct	Quantidade de palavras faladas na música.

## Tratamento de Dados

Linha	track_name	artist_name	f0_
1	SNAP	Rosa Linn	2
2	About Damn Time	Lizzo	2
3	Take My Breath	The Weeknd	2
4	SPIT IN MY FACE!	ThxSoMch	2

Durante a verificação da qualidade dos dados foram encontrados:

- **Valores duplicados:** Na tabela track\_in\_spotify, foram encontradas 4 duplicatas baseadas no campo track\_id. Os registros duplicados foram removidos para garantir a unicidade da chave primária. IDs excluídos: 8173823, 3814670, 1119309 e 7173596.
- **Valores ausentes (nulos):** Na tabela track\_technical\_info, a coluna key apresentava 95 valores nulos. Como a tonalidade musical nem sempre está disponível, esses valores foram tratados como “Ausente”, mantendo a observação na análise.
- **Ajustes de estrutura:** Na tabela track\_in\_spotify, os campos released\_year, released\_month e released\_day foram combinados em uma única coluna release\_date no formato de data padrão.
- **Foco estratégico nas plataformas:** Na tabela track\_in\_competition, a coluna in\_shazam\_charts foi descartada para concentrar a análise apenas nas plataformas concorrentes diretas: Apple Music e Deezer.

Criação de uma tabela única chamada track\_total\_info. Com as seguintes colunas:

track\_id, track\_name, artist\_name, artist\_count, release\_date (concatenação de released\_year, released\_month e released\_day), bpm, key\_note, mode, danceability\_pct, valence\_pct, energy\_pct, acousticness\_pct, instrumentalness\_pct, liveness\_pct, speechiness\_pct, in\_spotify\_playlists, in\_spotify\_charts, in\_apple\_playlists, in\_apple\_charts, in\_deezer\_playlists, in\_deezer\_charts, streams e total\_playlists (soma das playlists nas 3 plataformas).

## Classificação por Quartis

Para facilitar a análise das variáveis numéricas e permitir comparações mais estratégicas, criei uma tabela auxiliar com a distribuição dos **quartis** (Q1, Q2, Q3, Q4) e a **classificação** de cada variável relevante para as hipóteses.

As variáveis na tabela auxiliar são:

- bpm (batidas por minuto)
- streams (execuções no Spotify)
- total\_playlists (soma das playlists nas 3 plataformas)
- danceability\_pct, energy\_pct, valence\_pct, speechiness\_pct, acousticness\_pct, instrumentalness\_pct, liveness\_pct (percentuais relacionados às características das músicas)

Para cada variável, fiz a segmentação com base nos quartis, classificando os dados como baixo, médio ou alto:

**Quartil 1:** classificados como **Baixo**.

**Quartis 2 e 3:** classificados como **Médio**.

**Quartil 4:** classificados como **Alto**.

Para classificar as variáveis em níveis de desempenho (Baixo, Médio, Alto), utilizei a função NTILE(4) no SQL.

## Análise Exploratória de Dados (EDA)

A partir da tabela unificada track\_total\_info, foi iniciada a EDA no Power BI com o objetivo de entender a distribuição dos dados, identificar padrões, relações entre variáveis e possíveis outliers.

### Variáveis Categóricas

Os artistas aparecem mais de uma vez na base, não apenas com suas diferentes músicas, mas também em parcerias com outros artistas. De maneira individual os artistas com mais músicas são: **Taylor Swift** (34) e **The Weeknd** (21).

O tom musical mais comum entre as músicas é: **Dó Sustenido** (C#), sendo 16% da base. Também é o que tem maior número de streams. O modo mais frequente é **maior** com 58% da base e também com mais streams com 60%.

Dois artistas que têm mais presença em playlists no **Spotify** são The Weeknd e Taylor Swift. As músicas com maior presença na plataforma são "Blinding Lights" e "Shape of You".

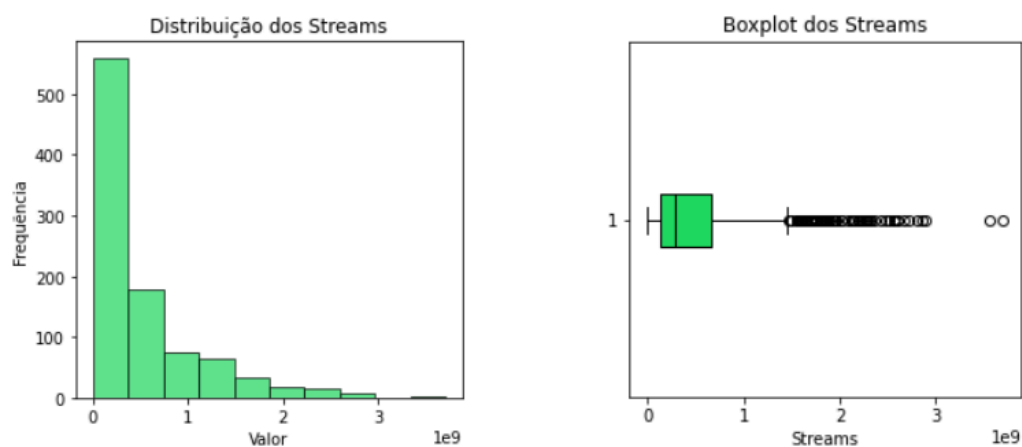
Já no **Deezer**, os artistas mais incluídos em playlists são Ed Sheeran e The Weeknd. As músicas com maior presença são "Shape of You" e "Dance Monkey".

No **Apple Music**, os artistas com mais playlists são Taylor Swift, Harry Styles e The Weeknd. As músicas mais incluídas são "Blinding Lights" e "Dance Monkey".

## Variáveis Numéricas

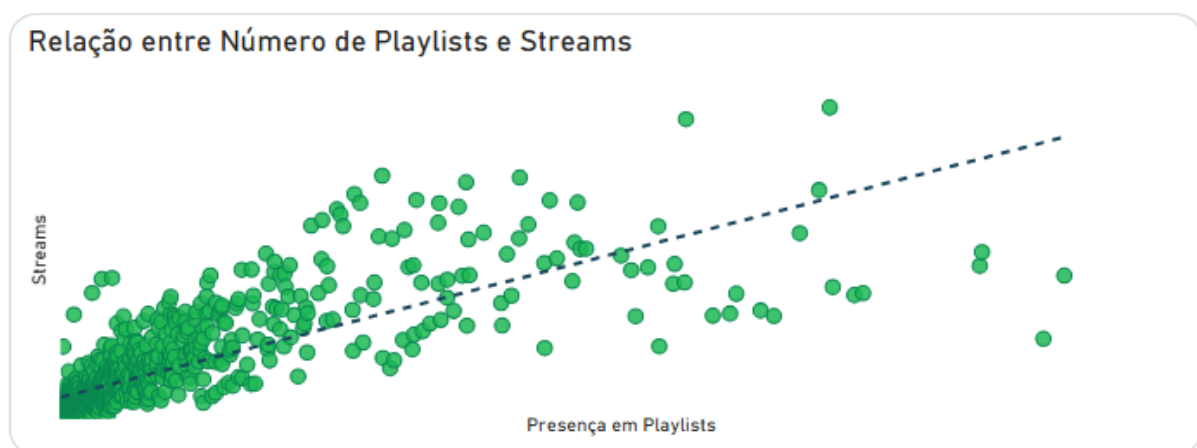
As variáveis numéricas associadas à popularidade das músicas foram **streams** e **total\_playlists** (soma das playlists em três plataformas: Deezer, Apple Music e Spotify).

A análise indicou que os **streams** têm uma **distribuição assimétrica**, com uma grande concentração em valores menores e alguns casos extremos de valores muito altos. O histograma apresentou distribuição do tipo **cauda longa**, onde uma minoria de músicas acumulam a maior parte dos streams. A presença de outliers, casos extremos, indica que algumas músicas tiveram desempenho excepcional.



A distribuição de **total\_playlists** segue um padrão similar ao de **streams**, com a maior parte das músicas listadas em poucas playlists e alguns casos de músicas incluídas em muitas playlists, indicando alta popularidade. O que indica que a presença da música em playlists pode ter um impacto significativo no **número de streams** de uma música.

Com um gráfico de dispersão foi possível observar a correlação entre número de streams e presença da música em playlists.



A base tem **muitos outliers**, porque algumas músicas fazem muito sucesso (mais streams) enquanto outras têm poucos streams. Isso cria uma distribuição bastante **assimétrica**, com **valores muito distantes da média**, um sinal clássico de que os dados **não são normais**.

**Dados normais:** seguem uma distribuição em forma de sino (curva normal), com a maioria dos valores próximos da média. **Dados não normais:** apresentam assimetrias, picos ou caudas alongadas, sem seguir a distribuição simétrica da curva normal.

A notação científica que aparece nos gráficos ( $1e9$ ) é por conta dos valores de streams que chegam na casa dos bilhões.  $1 \times 10^9$ , ou seja, **1 bilhão (1.000.000.000)**.

- A média de streams (513.794.564) é muito maior do que a mediana (288.101.651)
- O desvio padrão é alto (567.443.939)
- O histograma de streams tem uma cauda longa para a direita
- O boxplot mostra muitos pontos isolados acima dos limites superiores
- O valor mínimo de streams é 2762, máximo é 3.703.895.074

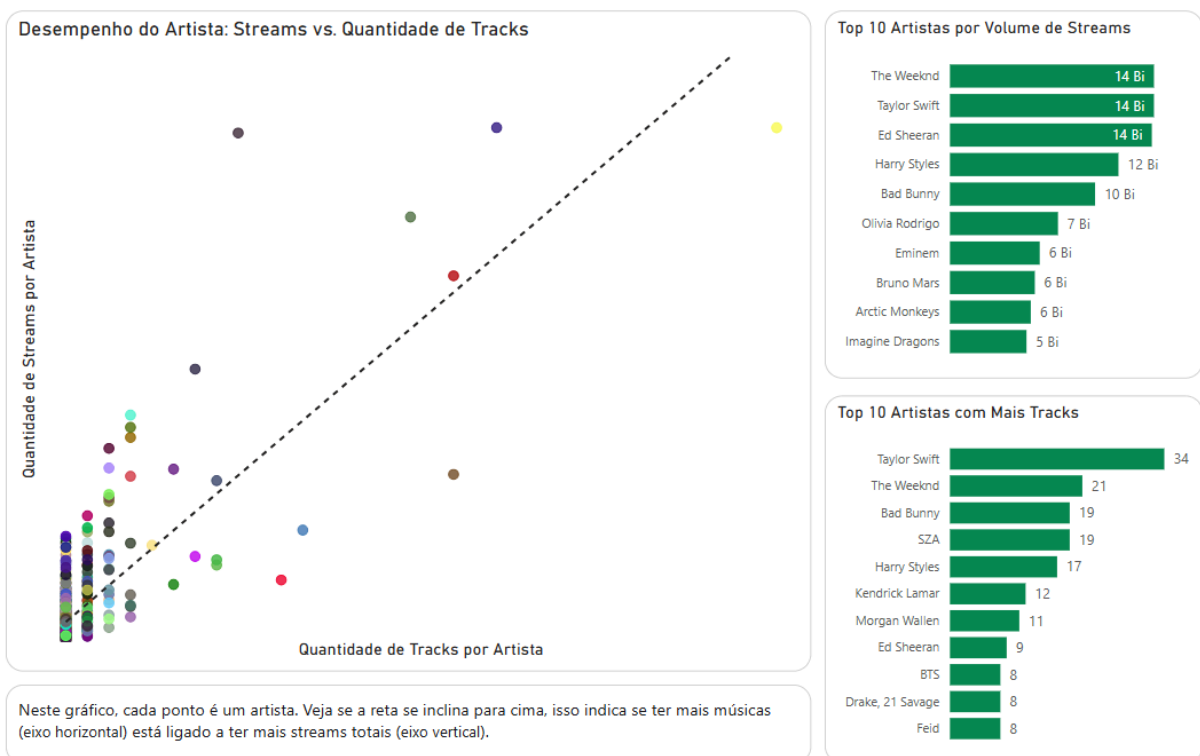
## Análise de Correlação

Além da distribuição das variáveis, também foi realizada uma **análise de correlação** para entender melhor como as variáveis associadas à popularidade das músicas (streams e playlists) se relacionam com outras variáveis.

- A **correlação entre streams e bpm** é muito baixa (-0.0019), indicando que o tempo da música não parece ter uma relação linear significativa com o número de streams.
- A **correlação entre streams e danceability** também é fraca (-0.1046), sugerindo que a dançabilidade das músicas não tem um impacto claro sobre a popularidade medida pelos streams.
- A **correlação entre streams e valence** (positividade da música) apresenta uma correlação ligeiramente negativa com os streams (-0.0426), o que sugere que músicas mais positivas não têm um impacto significativo na quantidade de streams.
- A **correlação entre streams e energia** das músicas tem uma correlação ainda mais fraca com os streams (-0.0262), sugerindo que a intensidade da música não está diretamente relacionada ao número de streams.
- A **correlação entre streams e acousticness** (acústica) e **instrumentalness** (instrumentalidade) apresentam correlações muito fracas e negativas com streams (-0.0043 e -0.0440, respectivamente), indicando que essas características não têm uma influência clara na popularidade.

- A **correlação entre streams e liveness** (ao vivo) e **speechiness** (percentual de fala) também não mostram uma correlação significativa com os streams, com valores de -0.0439 e -0.1120, respectivamente.
- A **correlação entre streams e total\_playlists** é bastante forte (0.7832), indicando uma **forte relação positiva**. Ou seja, músicas que aparecem em **mais playlists** tendem a gerar um **maior número de streams**.

Para validar a hipótese de que **artistas com um maior número de músicas** no Spotify também têm mais streams, calculei a correlação entre o número de faixas de cada artista e o total de streams gerados. Os resultados mostraram uma correlação forte e positiva de 0,78.

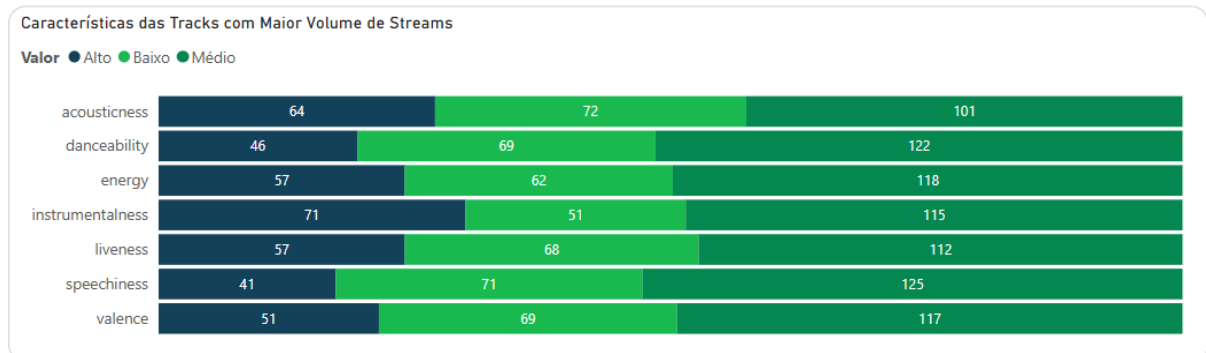


Isso indica que, de maneira geral, **artistas com mais músicas no Spotify tendem a acumular mais streams**, refletindo uma tendência significativa onde o volume de faixas disponíveis contribui para o aumento no número de streams. A relação não é perfeita, mas a correlação é suficientemente forte para validar a hipótese. Dessa forma, a hipótese foi confirmada.

### *Análise das Características das Músicas com Alta Quantidade de Streams*

Após a análise inicial das correlações entre as características das músicas (como danceability, energy, etc.) e o número de streams, percebi que não há uma correlação significativa entre essas variáveis e o sucesso de uma música (em termos de streams).

No entanto, para explorar de maneira mais aprofundada as possíveis influências, decidi fazer uma segmentação adicional, **filtrando as músicas com uma quantidade alta de streams** e analisando as características musicais dessas faixas.



Utilizando a segmentação por quartis, foquei no **quartil superior das músicas com mais streams**, ou seja, as músicas que se destacam no número de **execuções no Spotify**. Em seguida, criei gráficos de barra para cada uma das características da música, com o objetivo de entender **como as características se distribuem entre as músicas mais populares**.

A análise revelou que, apesar de as músicas com maior número de streams serem as mais populares, todas as características analisadas se posicionaram **no intervalo médio** (não em extremos como alto ou baixo) quando filtradas pelas músicas com mais streams.

Ou seja, **não há uma característica específica** que se destaque como um fator decisivo para o sucesso de uma música em termos de número de streams, mas sim uma **combinação de características musicais em níveis médios**.

## Variáveis Numéricas Ordinais

As variáveis **in\_spotify\_charts**, **in\_apple\_charts** e **in\_deezer\_charts** indicam a **presença e a posição** da música nas paradas do Spotify, Apple Music e Deezer, respectivamente. Os valores iguais a 0 representam ausência nas paradas, enquanto demais números indicam a posição ocupada pela música nos rankings.

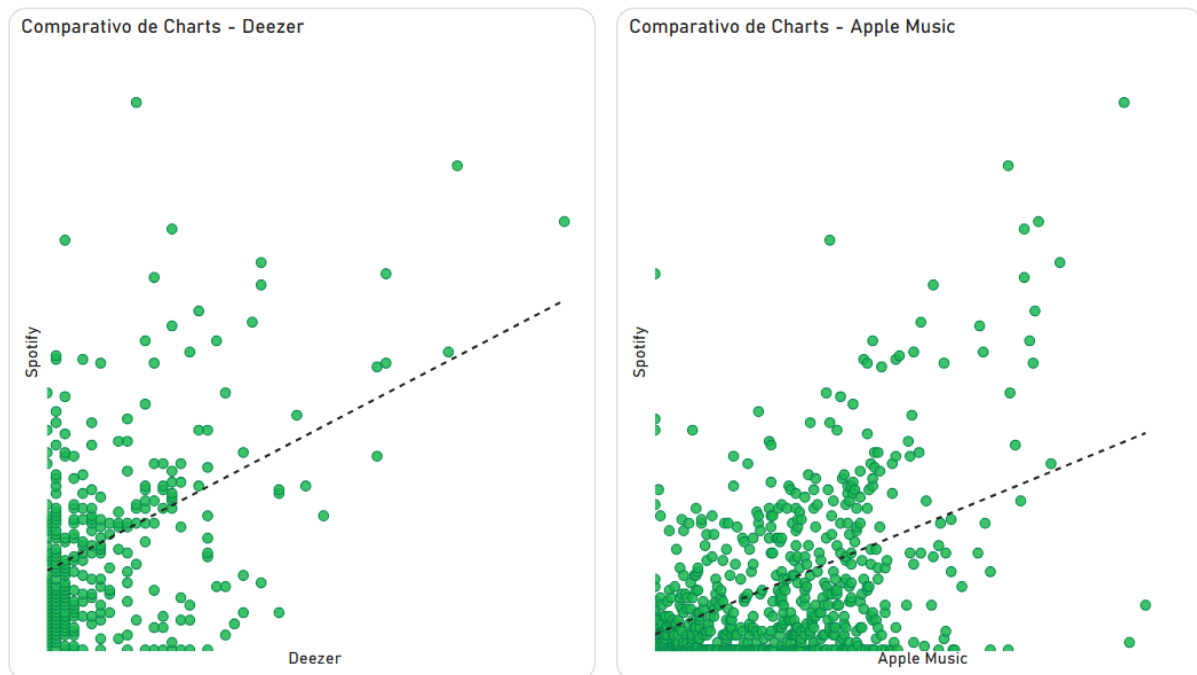
Para validar a hipótese de que músicas populares no ranking do Spotify também se destacam em outras plataformas como a Deezer e a Apple Music, calculei a correlação entre as posições nas paradas dessas plataformas.

Os resultados demonstraram uma correlação moderada positiva entre:

- Spotify e Apple Music (0,55)
- Spotify e Deezer (0,60)



Isso indica que há uma **tendência de que músicas populares em uma plataforma também estejam bem posicionadas em outras**, ainda que a relação não seja perfeita. A hipótese, portanto, foi confirmada.



## Validação das Hipóteses

**Hipótese:** *Músicas com BPM (Batidas Por Minuto) mais altos fazem mais sucesso em termos de número de streams no Spotify.*

Ambas as análises (Pearson e Spearman) mostram que não há uma **correlação positiva** entre o BPM e o número de streams no Spotify. A correlação de Pearson foi -0.0019 e a correlação de Spearman foi 0.0033, ambas muito próximas de zero.

Além disso, os valores p foram 0.9539 para Pearson e 0.9179 para Spearman, ambos muito elevados e bem maiores que 0.05, indicando que não há uma relação forte ou significativa entre essas variáveis.

**Resultado:** A análise de correlação demonstrou que **não existe relação** entre o BPM e a quantidade de streams. Dessa forma, **a hipótese foi refutada**, indicando que o ritmo mais acelerado de uma música não garante seu sucesso na plataforma.

**Hipótese:** *As músicas mais populares no ranking do Spotify também possuem um comportamento semelhante em outras plataformas, como a Deezer.*

Ambas as análises (Pearson e Spearman) indicam uma **correlação positiva** e estatisticamente significativa entre as posições das músicas no Spotify e nas outras

plataformas. A correlação de Pearson foi de 0.6049 entre Spotify e Deezer, e 0.5524 entre Spotify e Apple Music. Já a correlação de Spearman foi de 0.5916 para Deezer e 0.5217 para Apple Music.

Além disso, os valores p foram extremamente baixos, menores que 0,001, muito inferiores ao limite de 0,05, indicando uma forte relação significativa entre as variáveis.

**Resultado:** A análise da correlação **confirmou a hipótese**. Existe uma **correlação positiva** entre as posições das músicas no ranking do Spotify e nas outras plataformas (Apple Music e Deezer). Isso indica que, de forma geral, músicas populares em uma plataforma também tendem a ocupar boas posições em outras, ainda que a relação não seja perfeita.

**Hipótese:** *A presença de uma música em um maior número de playlists está correlacionada com um maior número de streams.*

Ambas as análises (Pearson e Spearman) confirmaram uma **correlação positiva e forte** entre o número de playlists em que uma música está presente e o número de streams. A correlação de Pearson foi 0.7832 e a correlação de Spearman foi 0.8325, ambas indicando uma relação significativa. Os valores p foram extremamente baixos, menores que 0,0001, o que sugere que a relação observada é **altamente significativa** e não é devido ao acaso.

**Resultado:** A hipótese foi **confirmada**. A análise de correlação indicou uma **relação positiva** entre o número de playlists em que a música aparece e sua quantidade de streams. Isso sugere que músicas adicionadas a mais playlists tendem a alcançar um público maior, o que contribui para um desempenho melhor em termos de popularidade na plataforma.

**Hipótese:** *Artistas com um maior número de músicas no Spotify têm mais streams.*

Ambas as análises (Pearson e Spearman) mostraram que existe uma **correlação positiva** entre o número de músicas de um artista e o número de streams. A correlação de Pearson foi de 0.779, indicando uma relação moderada a forte, enquanto a correlação de Spearman foi de 0.448, indicando uma correlação moderada, embora não tão forte. Ambos os valores p são extremamente baixos, o que indica que a correlação é estatisticamente significativa.

**Resultado:** A análise da correlação entre o número de músicas de um artista (quantidade de faixas) e o número de streams mostrou uma **correlação positiva**. Isso significa que, de forma geral, artistas que possuem mais faixas no Spotify tendem a ter mais streams. Dessa forma, a hipótese foi **confirmada**.

**Hipótese:** *As características da música influenciam o sucesso em termos de número de streams no Spotify.*

**Resultado:** A análise de correlação entre as principais características musicais e o número de streams no Spotify mostrou que **não há correlação entre essas variáveis**. Todas as correlações observadas foram muito próximas de zero. Esses resultados indicam que **nenhuma dessas características isoladamente** parece impactar o número de streams de forma relevante. Dessa forma, a hipótese foi refutada.

Porém, analisando as **características por meio dos quartis**, foi possível identificar uma **predominância de equilíbrio** nas características das músicas mais populares, com valores geralmente concentrados em **faixas médias**. Isso pode sugerir que, **embora não exista correlação linear**, há uma preferência do público por músicas com **características balanceadas**, sem extremos.

Hipótese	Resultado
BPM mais alto resulta em mais streams	Refutada
Popularidade no Spotify correlaciona com outras plataformas	Confirmada
Músicas em mais playlists têm mais streams	Confirmada
Artistas com mais músicas têm mais streams	Confirmada
Características musicais influenciam o sucesso (mais streams)	Refutada

## Teste de Significância

**Hipótese:** *A presença de uma música em um maior número de playlists está correlacionada com um maior número de streams. Esse padrão também acontece com as músicas lançadas em 2023?*

**Tipo de teste:** Correlação de Pearson e Spearman

A **correlação de Pearson** (0.7846) indica uma relação **forte** entre as variáveis.

A **correlação de Spearman** (0.6539) indica uma relação **moderada**, mas ainda significativa.

Mesmo para músicas lançadas em 2023 há uma associação significativa entre o número de playlists e o número de streams.

A diferença entre os dois testes é que o **Pearson** verifica se as variáveis têm uma relação **constante e proporcional**, ou seja, se aumentam ou diminuem de forma linear, **como uma linha reta**.

Já o **Spearman** analisa se as variáveis tendem a aumentar ou diminuir juntas, mas **não necessariamente de forma constante ou linear**, permitindo que a relação entre elas seja mais flexível, **como uma curva**.

Como o teste de **Pearson** é sensível a outliers e pressupõe linearidade dos dados, ele pode acabar exagerando a força da correlação se houver valores extremos. Já o teste de **Spearman**, por ser baseado em postos e não nos valores brutos, tende a suavizar o impacto dos outliers, resultando em uma correlação aparentemente mais fraca.

Como temos **outliers**, Spearman reflete melhor a **verdadeira relação entre as variáveis**, enquanto Pearson pode ser um pouco distorcido por esses valores extremos.

**Hipótese:** *Músicas com colaborações (feat) têm, em média, mais streams do que músicas solo?*

**Tipo de teste:** Mann-Whitney

**Hipótese nula ( $H_0$ ):** as distribuições de streams são iguais entre músicas com e sem colaboração

**Hipótese alternativa ( $H_1$ ):** as distribuições são diferentes

Embora o gráfico de barras mostre que faixas solo somam mais streams, isso acontece porque há mais faixas desse tipo na base. Porém, ao aplicar o teste de Mann-Whitney (mais adequado para esse tipo de dado com outliers e distribuição não normal), foi possível verificar que **faixas com colaborações têm, em média, mais streams que faixas solo**.

O teste retornou uma estatística  $U = 88.309,0$  e um p-valor muito baixo ( **$7.75e-06$** ), indicando que a diferença observada é estatisticamente significativa.

Esse teste é **não paramétrico**, então ele **não assume normalidade** nem é tão afetado por outliers. É ótimo para comparar medianas entre dois grupos.

Testes não paramétricos são usados quando os dados **não seguem uma distribuição normal** ou quando **há outliers** que poderiam distorcer os resultados de testes tradicionais. Em vez de comparar médias, eles analisam **medianas** e distribuições.

O **teste Mann-Whitney**, que compara duas amostras independentes e verifica se uma tende a ter valores maiores que a outra.

Outro é o **teste de Wilcoxon**, que funciona para amostras pareadas. Esses testes são muito úteis em pesquisas onde os dados são assimétricos ou possuem variabilidade alta.

Hipótese	Resultado
Músicas de 2023 em mais playlists têm mais streams	Confirmada
Músicas com colaborações têm mais streams	Confirmada

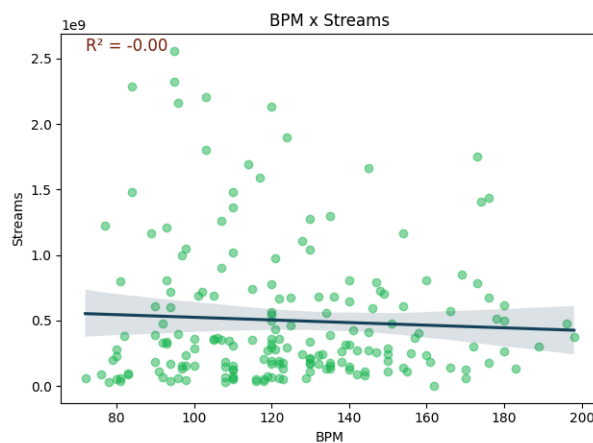
## Regressão Linear

**Hipótese:** *Músicas com BPM mais altos fazem mais sucesso em termos de streams no Spotify.*

- ❖ **Variável independente (X):** BPM
- ❖ **Variável dependente (Y):** Streams

Os dados foram divididos em treino (80%) e teste (20%) para aplicar o modelo de regressão linear.

**Avaliação do modelo:** O valor de  $R^2$  foi **-0,005**, indicando que o modelo não consegue explicar a variação da variável dependente com base na variável independente. Também foram aplicadas as correlações de Pearson e Spearman, ambas indicando ausência de correlação entre as variáveis.



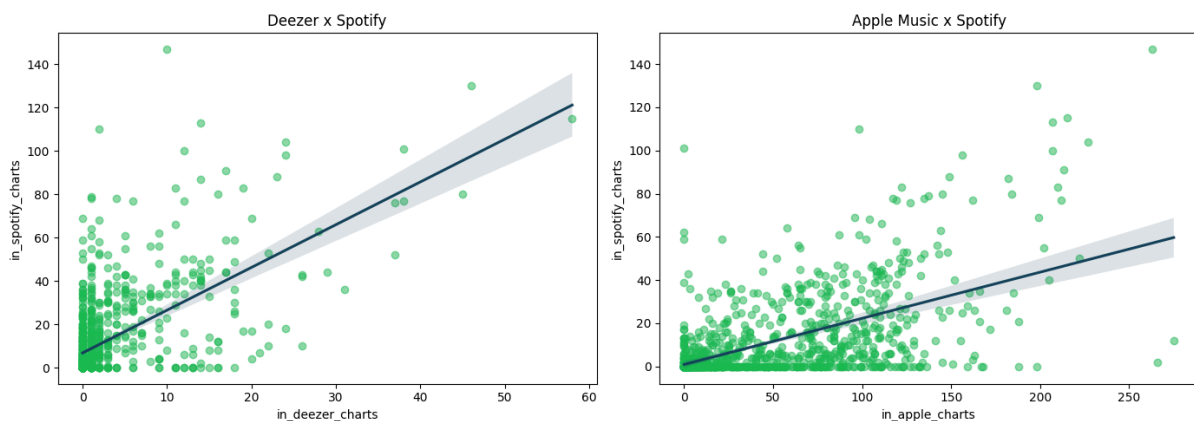
**Conclusão:** Não foi identificada uma relação linear entre o BPM e o número de streams. Então, a hipótese foi **refutada** com base nos testes estatísticos e na análise do modelo de regressão.

**Hipótese:** *As músicas mais populares no ranking do Spotify também possuem um comportamento semelhante em outras plataformas como Deezer.*

- ❖ **Variáveis independentes (X):** Posições nas paradas do Deezer e Apple Music
- ❖ **Variável dependente (Y):** Posição no Spotify

Os dados foram divididos em treino (80%) e teste (20%) para aplicar o modelo de regressão linear múltipla.

**Avaliação do modelo:** O valor de  $R^2$  foi 0.486, indicando que o modelo explica aproximadamente 48.6% da variação na posição no Spotify com base nas posições nas paradas do Deezer e Apple Music. Ambos os p-valores para as variáveis independentes (Deezer e Apple Music) foram significativamente menores que 0.05, indicando que há uma relação estatisticamente significativa entre as variáveis independentes e a variável dependente.



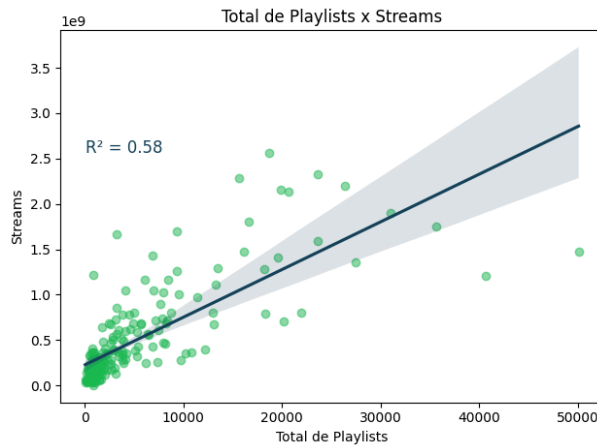
**Conclusão:** Foi identificada uma relação positiva e significativa entre as posições nas paradas do Deezer e Apple Music e a posição no Spotify. A hipótese de que a posição nas paradas do Deezer e Apple Music influencia a posição no Spotify foi **confirmada** com base nos resultados da regressão linear múltipla.

**Hipótese:** *A presença de uma música em um maior número de playlists está relacionada com um maior número de streams.*

- ❖ **Variável independente (X):** Total de Playlists
- ❖ **Variável dependente (Y):** Streams

Os dados foram divididos em treino (80%) e teste (20%) para aplicação do modelo de regressão linear.

**Avaliação do modelo:** O valor de  $R^2$  foi **0.58**, indicando que 58% da variação na quantidade de streams pode ser explicada pela presença da música em playlists.



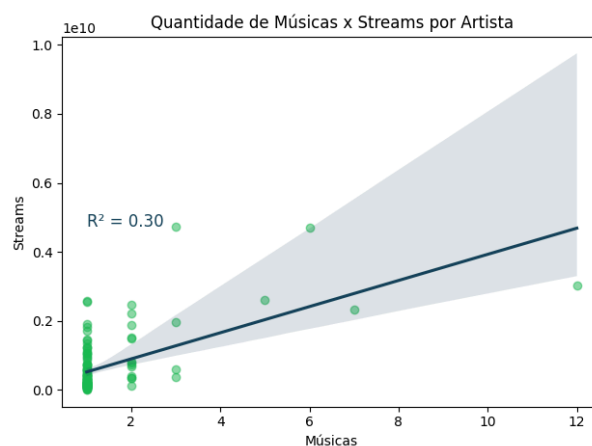
**Conclusão:** Foi identificada uma relação linear positiva entre o número de playlists e os streams. A hipótese foi confirmada com base no desempenho do modelo de regressão e na visualização da tendência dos dados.

**Hipótese:** *Artistas com maior número de músicas no Spotify têm mais streams.*

- ❖ **Variável independente (X):** Quantidade de músicas por artista
- ❖ **Variável dependente (Y):** Total de streams por artista

Os dados foram agregados por artista, contabilizando o total de faixas e a soma dos streams. Após a preparação, foi aplicado um modelo de regressão linear simples com divisão entre treino (80%) e teste (20%).

**Avaliação do modelo:** O valor de  $R^2$  foi aproximadamente 0.30, o que indica uma correlação fraca, mas presente. Isso significa que artistas com mais músicas tendem a ter mais streams, embora essa variável por si só não explique a maior parte da variação nos dados.



**Conclusão:** Apesar de a **regressão linear não ser confiável** para explicar a variação nos streams com base apenas no número de músicas, os **testes de correlação confirmam** a

existência de uma relação positiva entre as variáveis. Assim, a hipótese foi confirmada com base na análise de correlação, e não na regressão.

**Hipótese:** *As características da música influenciam o sucesso em termos de streams no Spotify.*

- ❖ **Variáveis independentes (X):** danceability\_pct, valence\_pct, energy\_pct, acousticness\_pct, instrumentalness\_pct, liveness\_pct, speechiness\_pct
- ❖ **Variável dependente (Y):** Streams

Foi aplicada uma **regressão linear múltipla** para investigar a relação entre as características da música e o número de streams.

O modelo apresentou um valor de  $R^2$  de 0.029 (ou 2.9%), indicando que o conjunto dessas características musicais **explica apenas cerca de 3% da variação nos streams**. Esse valor é muito baixo e sugere que, analisadas conjuntamente neste modelo linear, as **características técnicas não são boas preditoras** do sucesso de uma música em termos de streams.

A maioria das características individuais não apresentou relação estatisticamente significativa com os streams ( $p$ -valores  $> 0.05$ ), incluindo valence\_pct, energy\_pct, acousticness\_pct, instrumentalness\_pct e liveness\_pct.

Apenas danceability\_pct ( $p=0.005$ ) e speechiness\_pct ( $p=0.002$ ) mostraram uma relação negativa estatisticamente significativa, sugerindo que, mantendo as outras características constantes, **músicas mais dançantes ou com mais palavras faladas** podem estar associadas a **menos streams neste conjunto de dados**.

**Conclusão:** Com base neste modelo, a hipótese de que essas características técnicas influenciam significativamente o sucesso em streams **não foi confirmada**. O sucesso provavelmente depende muito mais de outros fatores não incluídos no modelo (marketing, playlists, artista, etc.) ou de interações mais complexas. A baixa explicação do modelo indica que **outras variáveis ou uma análise mais complexa são necessárias** para entender melhor os determinantes do sucesso em streams.

### **Resumo da Regressão Linear Múltipla (OLS)**

Estes foram os principais resultados do modelo de regressão linear múltipla aplicado para investigar a relação entre características técnicas das músicas e o número de streams no Spotify. (biblioteca statsmodels)



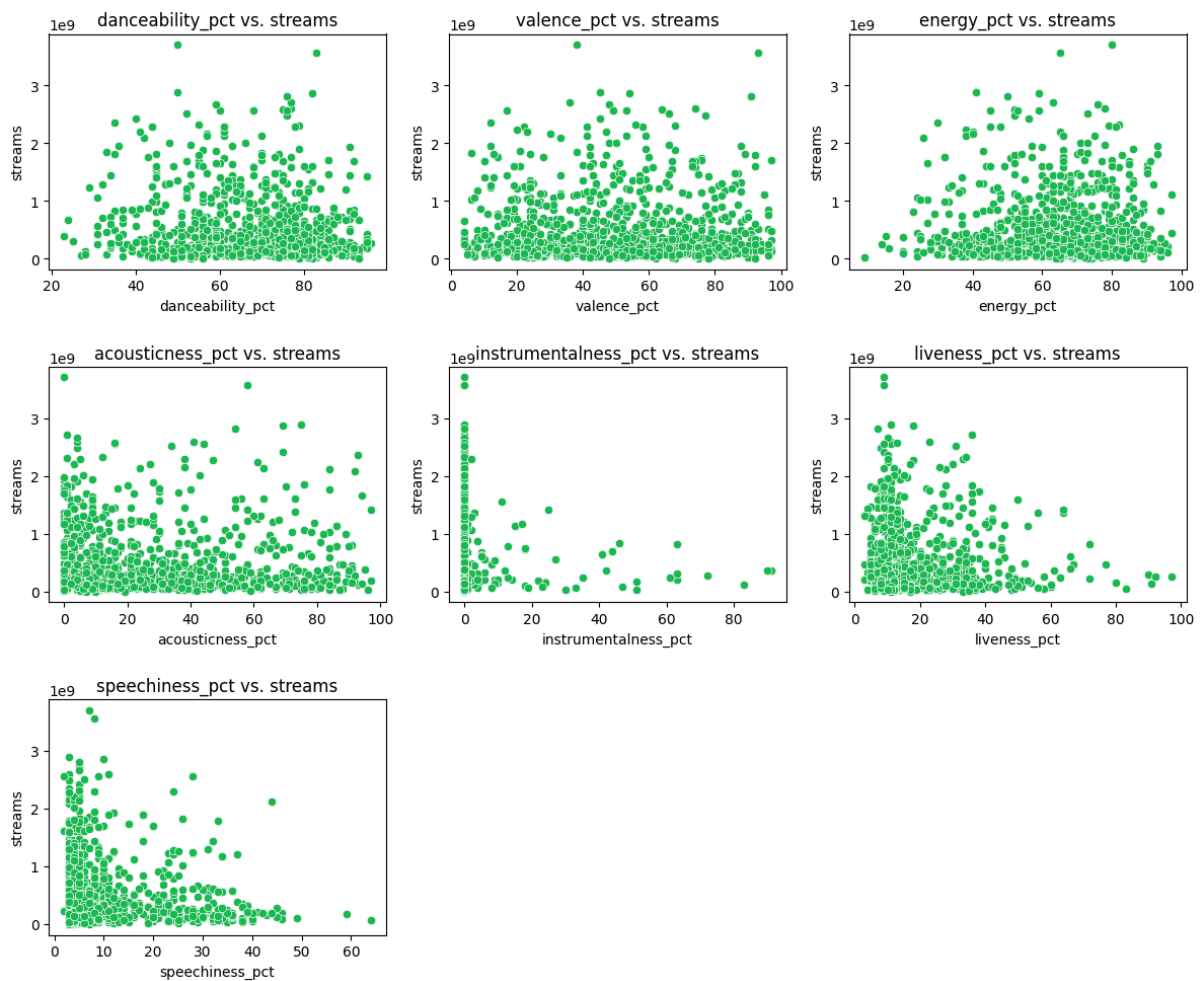
Métrica	Valor	Interpretação
R <sup>2</sup>	0,029	As variáveis explicam apenas 2,9% da variação nos streams
R <sup>2</sup> ajustado	0,021	Ajustado pelo número de variáveis, ainda baixa capacidade de explicar
Estatística F	3,974	Mede se o modelo como um todo é significativo
Prob (F-statistic)	0,000272	O modelo tem significância estatística geral ( $p < 0,05$ )
Danceability	-4.0 mi	Relação negativa e significativa com streams ( $p = 0.005$ )
Speechiness	-5.8 mi	Relação negativa e significativa com streams ( $p = 0.002$ )
Outras variáveis	$p > 0,05$	Não apresentaram relação estatisticamente significativa com os streams
Durbin-Watson	1,773	Valor próximo de 2, indica ausência de autocorrelação forte nos resíduos

A análise visual dos gráficos de dispersão corrobora os resultados do modelo de regressão linear múltipla (OLS). A **ausência de tendências lineares fortes** na maioria dos gráficos explica o baixo R<sup>2</sup> geral do modelo (0.029), indicando que essas características, combinadas linearmente, têm **baixo poder preditivo** sobre os streams.

Especificamente, o **padrão visual negativo** observado para speechiness e, de forma mais sutil, para danceability, isso mostra os **coeficientes negativos** e **estatisticamente significativos** encontrados na regressão para essas variáveis.

Da mesma forma, a dispersão sem padrão claro para valence, energy e acousticness corresponde aos seus p-valores não significativos no modelo.

O gráfico de instrumentalness mostra uma **forte relação visual negativa**, que o modelo linear OLS capturou apenas no limite da significância ( $p=0.054$ ), possivelmente devido a não-linearidades na relação.



Em conjunto, tanto a análise visual quanto a estatística reforçam a conclusão de que as características técnicas isoladas não são os principais impulsionadores do sucesso em streams, que provavelmente depende de uma gama mais ampla de fatores.

## Conclusão da Análise

- A presença de uma música em playlists está fortemente ligada ao seu número de streams, sendo um fator chave para visibilidade e popularidade, inclusive para lançamentos recentes (2023).
- Músicas com colaborações (feats) tendem a gerar mais streams do que músicas solo, sugerindo que parcerias são eficazes para ampliar o alcance.
- Músicas populares no Spotify geralmente performam bem também no Deezer e Apple Music, indicando uma sinergia entre os principais serviços de streaming.
- Artistas com maior presença no catálogo tendem a acumular mais streams, mostrando a relevância da consistência e do volume de lançamentos.

- As características técnicas da música, analisadas isoladamente, não mostraram forte correlação direta com o sucesso em streams. A análise sugere que um equilíbrio nessas características pode ser preferível a valores extremos.
- Fica claro que o sucesso de uma música é complexo e depende de uma combinação de fatores que vão além das características técnicas, incluindo estratégias de marketing, contexto cultural e engajamento do público.

## Recomendações

- ★ **Playlists:** Priorizar a inclusão estratégica das músicas em playlists relevantes (editoriais, de curadores, de usuários) como principal tática para aumentar streams e descobertas.
- ★ **Colaborações:** Encarar as parcerias (feats) como uma ferramenta valiosa para expandir a audiência e aumentar o potencial de engajamento de um novo artista.
- ★ **Estratégia Além de Acordes:** Reconhecer que o sucesso não vem só da música. Investir em marketing direcionado, promoção de engajamento social e compreensão do contexto cultural para maximizar o impacto.
- ★ **Abordagem Multiplataforma:** Desenvolver estratégias de lançamento e promoção que integram Spotify, Deezer e Apple Music, aproveitando a tendência de sucesso cruzado entre elas.
- ★ **Equilíbrio Sonoro:** Na produção musical, considerar que músicas com características técnicas balanceadas (nem muito altas, nem muito baixas) podem ter maior apelo, embora a autenticidade do artista deva prevalecer.
- ★ **Análise Contínua:** Para futuras análises, incluir dados de gênero musical e informações sobre campanhas de marketing para obter uma visão ainda mais completa dos fatores de sucesso.

## Links do Projeto:

[Dashboard no Power BI](#)

[Apresentação no Google Slides](#)

[Repositório no GitHub](#)

## Links de Estudo:

- ▶ [TESTE DE HIPÓTESE, P VALOR E NÍVEL DE SIGNIFICÂNCIA](#)
- ▶ [TESTE PARAMÉTRICO E NÃO PARAMÉTRICO - DIFERENÇAS E QUANDO USAR CADA UM](#)
- ▶ [Análise de Regressão Linear Múltipla - Definição](#)