



Projeto 2 - Hipóteses 🎵

Ficha Técnica:

1. Objetivo da análise

No cenário musical global, onde a indústria é extremamente competitiva e está em constante evolução, a capacidade de tomar decisões baseadas em dados tornou-se um ativo inestimável.

Nesse contexto, uma gravadora enfrenta o emocionante desafio de lançar um novo artista. Felizmente, ela possui uma ferramenta poderosa em seu arsenal: um extenso conjunto de dados do Spotify com informações sobre as músicas mais ouvidas em 2023.

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

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

Você deve validar (refutar ou confirmar) essas hipóteses através da análise de dados e fornecer recomendações estratégicas com base em suas descobertas. O objetivo principal desta análise é que a gravadora e o novo artista possam tomar decisões informadas que aumentem suas chances de alcançar o “sucesso”.

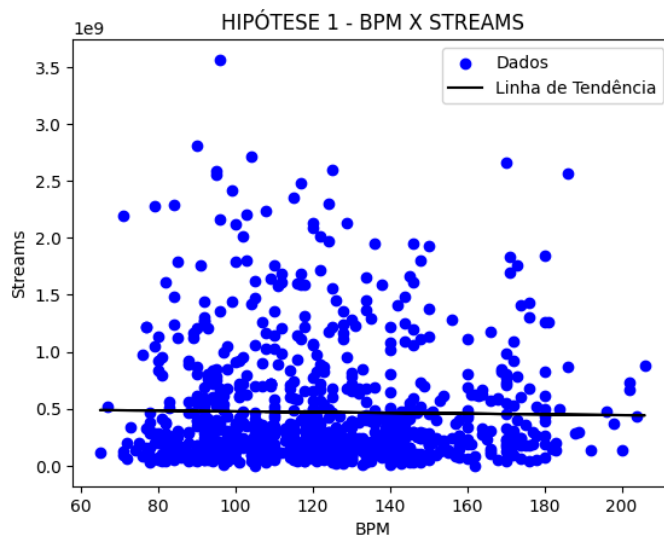
Análise Geral

Nossa base de dados tratada possui:

- 614 artistas de diversos ramos musicais
- 892 músicas
- 843 Playlists
- 425 bilhões de Streams

Hipóteses

1 - Músicas com BPM (Batidas Por Minuto) mais altas fazem mais sucesso em termos de número de streams no Spotify.



R: O gráfico intitulado "HIPÓTESE 1 - BPM X STREAMS" apresenta uma dispersão de pontos que representam a relação entre o número de batimentos por minuto (BPM) e o número de streams. Aqui estão algumas observações detalhadas sobre o gráfico:

Eixo X (BPM): Varia de aproximadamente 60 a 200 BPM.

Eixo Y (Streams): Varia de 0 a aproximadamente 3.5 bilhões de streams.

Pontos de Dados: Representados em azul, mostram a distribuição dos streams em relação aos BPMs.

Linha de Tendência: Uma linha preta horizontal atravessa o gráfico, indicando que não há uma tendência clara ou significativa que mostra um aumento ou diminuição no número de streams com o aumento dos BPMs.

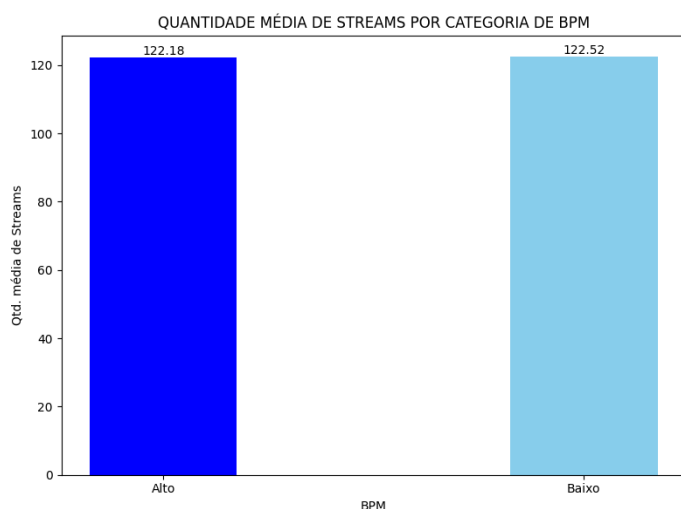
Análise:

A linha de tendência horizontal sugere que não há uma correlação forte entre o número de BPM e o número de streams. Isso significa que variações no BPM das músicas não parecem influenciar significativamente o número de streams que elas recebem.

A dispersão dos pontos é bastante ampla e não segue um padrão claro, reforçando a ideia de que outros fatores além do BPM podem ser mais determinantes para o número de streams.

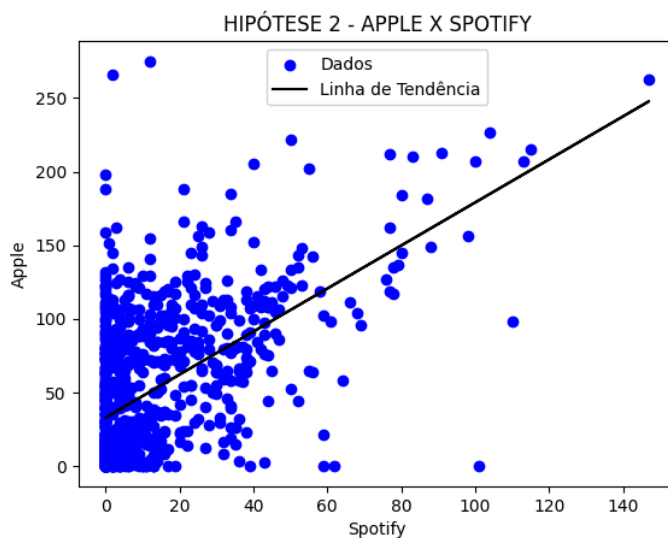
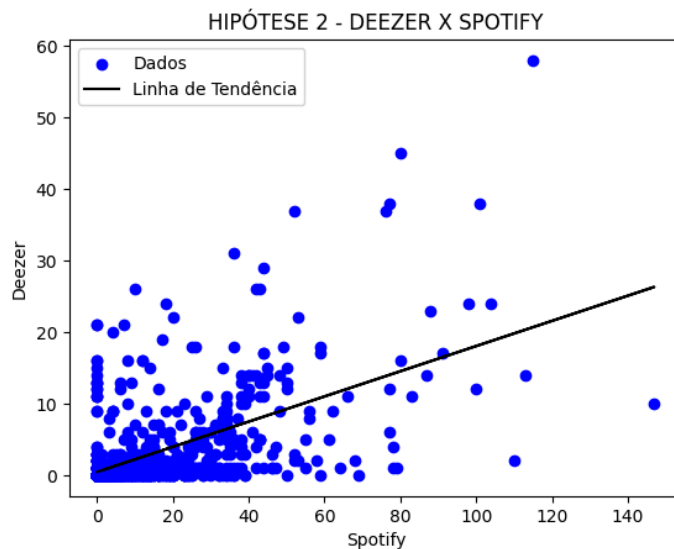
Conclusão:

A hipótese de que o BPM alto influencia diretamente o número de streams não é suportada por este gráfico, indicando que as preferências de streaming de música podem ser influenciadas por uma variedade de outros fatores, como gênero musical, popularidade do artista, qualidade da música, entre outros. **Hipótese refutada!**



No gráfico fica evidenciado que uma música com um ritmo mais acelerado, medido em beats por minuto (bpm), apresenta um valor médio de 122,18 streams. Por outro lado, uma música com um ritmo mais lento tem uma média de aproximadamente 122,52 streams, uma diferença de apenas 0,34. Isso representa uma variação de aproximadamente 0,28%, indicando que a influência do ritmo na quantidade de streams é mínima.

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



R: Os gráficos intitulado "HIPÓTESE 2 - DEEZER X SPOTIFY" e "HIPÓTESE 2 - APPLE X SPOTIFY" apresentam uma análise de correlação entre a popularidade das músicas no Spotify e no Deezer e Apple. Aqui estão algumas observações detalhadas sobre o gráfico:

No eixo x, temos a popularidade no Spotify.

No eixo y, a popularidade no Deezer e Apple.

Os pontos azuis representam os dados individuais de popularidade das músicas nas duas plataformas.

A linha de tendência preta no gráfico indica uma relação positiva entre as popularidades no Spotify e no Deezer.

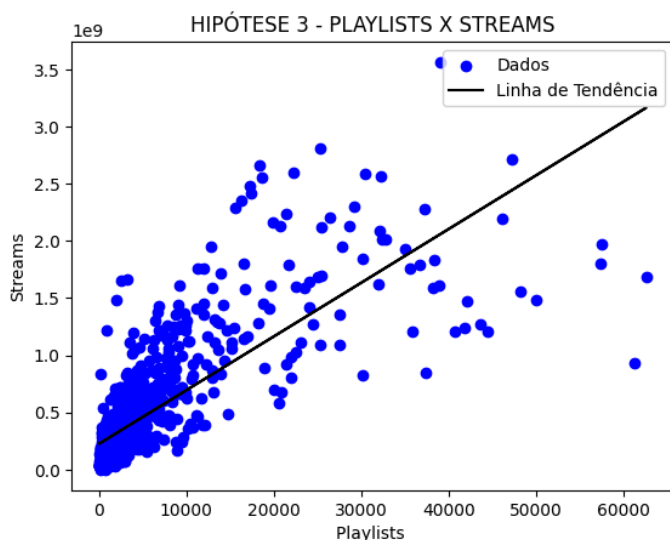
Análise:

Sugerindo que músicas populares no Spotify tendem a ter um desempenho semelhante no Deezer. No entanto, a dispersão dos pontos ao redor da linha de tendência mostra que essa relação não é perfeitamente linear ou extremamente forte, indicando variações no comportamento das músicas entre as duas plataformas.

Conclusão:

Essa análise corrobora a ideia de que, embora exista uma tendência de músicas populares no Spotify também serem populares no Deezer ou na Apple, existem exceções e variações que podem ser influenciadas por outros fatores específicos de cada plataforma ou pelo gosto dos usuários em diferentes regiões. **Hipótese confirmada!**

3 - A presença de uma música em um maior número de playlists está correlacionada com um maior número de streams.



R: A imagem exibe um gráfico de dispersão intitulado "HIPÓTESE 3 - PLAYLISTS X STREAMS". Ele traça a relação entre o número de playlists e o número de streams, com playlists no eixo x e streams no eixo y.

O eixo x varia de 0 a 60.000 playlists.

O eixo y varia de 0 a 3,5 Bilhões streams.

O gráfico inclui vários pontos azuis que representam pontos de dados individuais e uma linha de tendência preta indicando uma correlação positiva entre o número de playlists e o número de streams.

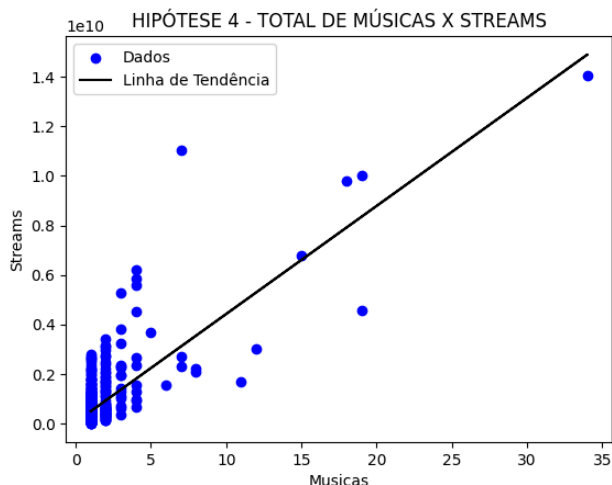
Análise:

Isso sugere que à medida que aumenta o número de playlists, o número de streams também tende a aumentar. O gráfico é utilizado para analisar visualmente os dados e apoiar a hipótese de que existe relação entre as duas variáveis.

Conclusão:

A análise da imagem indica que existe uma correlação positiva entre o número de playlists e o número de streams dos dados visualizados. Conforme análise de correlação, vimos que o número de playlists que a determinada música se encontra faz toda a diferença e influencia diretamente no número de streams, ou seja quanto mais playlists e música estiver, maior será o número de streams que ela terá. **Hipótese confirmada!**

4 - Artistas com um maior número de músicas no Spotify têm mais streams.



R: Análise do Gráfico: "HIPÓTESE 4 - TOTAL DE MÚSICAS X STREAMS", Este gráfico de dispersão analisa a relação entre o número total de músicas que um artista possui e o número total de streams que essas músicas receberam.

No Eixo X (Músicas): Representa o número total de músicas de cada artista.

No Eixo Y (Streams): Representa o número total de streams que essas músicas receberam, medidos em bilhões (1,4).



Os Pontos Azuis: Cada ponto representa um artista individual, mostrando a relação entre o número de músicas e o número total de streams. Na Distribuição dos Pontos a maioria dos artistas tem menos de 10 músicas e, conseqüentemente, menos de 0,4 bilhões de streams. Há alguns artistas com um número significativamente maior de músicas (até 35) e com streams que chegam a 1,4 bilhões.

A Linha de Tendência Preta: Indica a tendência geral dos dados, mostrando a direção da relação entre o número de músicas e o número de streams. Já na Linha de Tendência a ascendente indica uma correlação positiva entre o número de músicas e o número de streams. Isso sugere que, à medida que o número de músicas aumenta, o número total de streams também tende a aumentar.

Análise:

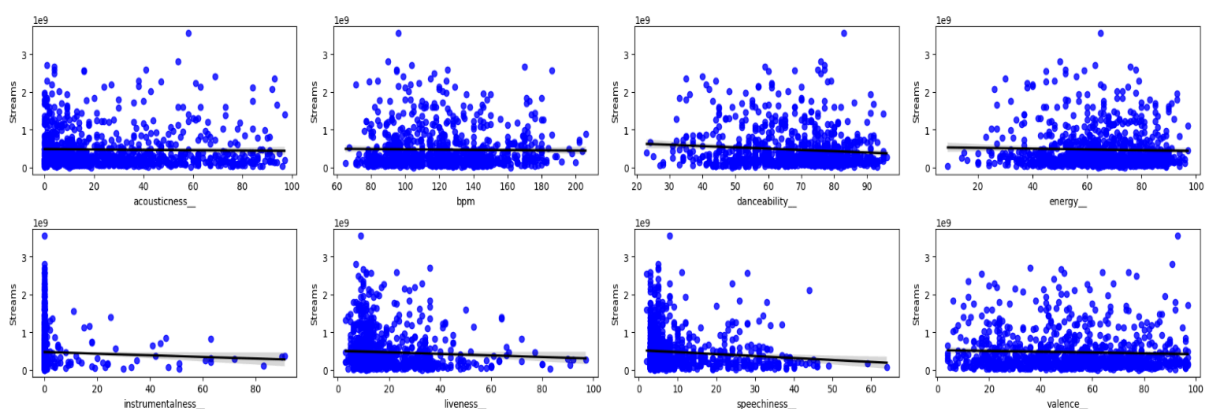
Na Dispersão dos Dados embora haja uma correlação positiva, a dispersão dos pontos sugere que a relação não é perfeitamente linear. Alguns artistas com um número baixo de músicas ainda conseguem um número relativamente alto de streams, e vice-versa.

Conclusão:

Existe uma correlação positiva entre o número de músicas e o número total de streams. Isso indica que artistas com mais músicas tendem a acumular mais streams no total. A estratégia de lançamento frequente de músicas pode ser benéfica para artistas que desejam aumentar seus streams. Artistas com um catálogo mais extenso têm mais chances de alcançar um público maior e gerar mais streams. **Hipótese confirmada!**

5 - As características da música influenciam o sucesso em termos de número de streams no Spotify.

HIPÓTESE 5 - STREAMS X CARACTERÍSTICAS



R: A imagem mostra uma série de gráficos de dispersão analisando a relação entre o número de streams (no eixo y) e várias características musicais (no eixo x). O título "HIPÓTESE 5 - STREAMS X



CARACTERÍSTICAS" sugere que esta análise está testando uma hipótese sobre como essas características influenciam o número de streams.

Aqui está uma análise detalhada de cada gráfico:

acousticness__ vs. Streams: Há uma concentração de pontos de dados em valores mais baixos de acousticness__, com alguns outliers. A linha de tendência sugere uma leve correlação negativa, indicando que valores mais altos de acousticness__ podem estar associados a um menor número de streams.

bpm vs. Streams: O gráfico de dispersão para bpm (batidas por minuto) mostra uma ampla gama de valores de bpm, com uma leve concentração em valores mais baixos. A linha de tendência é quase horizontal, indicando pouca ou nenhuma correlação entre bpm e o número de streams.

danceability__ vs. Streams: Este gráfico mostra uma distribuição semelhante, com uma leve linha de tendência horizontal, sugerindo pouca ou nenhuma correlação entre danceability__ e streams.

energy__ vs. Streams: Novamente, os pontos de dados estão amplamente espalhados com uma linha de tendência horizontal, indicando uma relação mínima entre energy__ e streams.

instrumentalness__ vs. Streams: Há uma concentração clara de pontos em valores baixos de instrumentalness__, com muito poucas faixas instrumentais recebendo um número significativo de streams. A linha de tendência sugere uma leve correlação negativa.

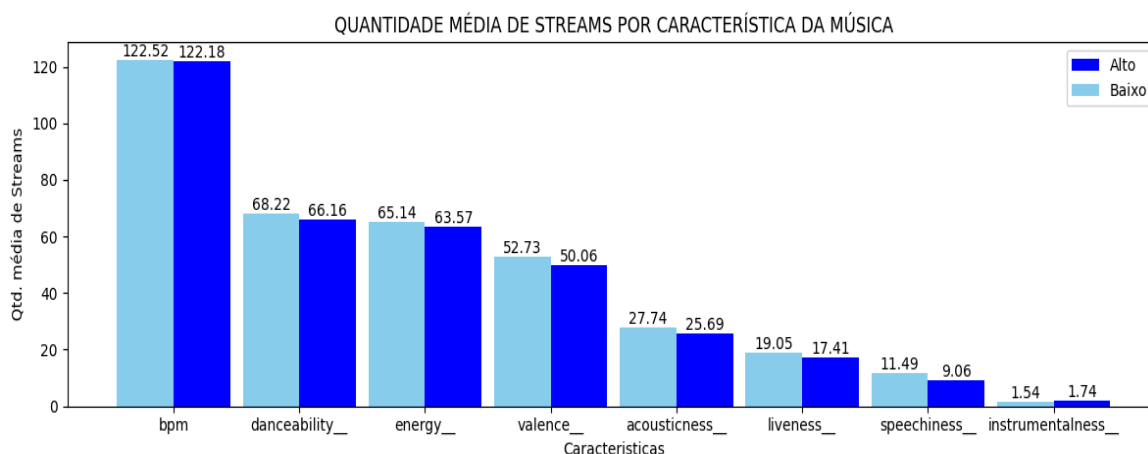
liveness__ vs. Streams: Os pontos de dados estão espalhados ao longo do espectro de liveness__, com uma linha de tendência quase horizontal, indicando nenhuma correlação forte entre liveness__ e streams.

speechiness__ vs. Streams: Este gráfico mostra uma densa concentração de pontos em valores mais baixos de speechiness__, com uma linha de tendência indicando uma leve correlação negativa.

valence__ vs. Streams: O gráfico de dispersão para valence__ também mostra uma distribuição de pontos de dados com uma linha de tendência quase plana, sugerindo pouca ou nenhuma correlação entre valence__ e streams.

Conclusão:

No geral, as linhas de tendência nesses gráficos de dispersão sugerem que há correlações fracas ou negligenciáveis entre o número de streams e essas características musicais individuais. Isso implica que outros fatores podem ser mais influentes na determinação da popularidade de uma música em termos de streams. **Hipótese refutada!**



Já na Análise do Gráfico: QUANTIDADE MÉDIA DE STREAMS POR CARACTERÍSTICA DA MÚSICA. O gráfico de barras compara a quantidade média de streams para diferentes características das músicas, divididas entre valores altos e baixos dessas características.

O Eixo X (Características): Representam diferentes características das músicas, como BPM, danceability, energy, valence, acousticness, liveness, speechiness, e instrumentalness.

O Eixo Y (Qtd. média de Streams): Representa a quantidade média de streams recebidos pelas músicas.

Nas Barras Azuis (Alto): Indicam a quantidade média de streams para músicas com valores altos dessas características.

Já nas Barras Azul Claro (Baixo): Indicam a quantidade média de streams para músicas com valores baixos dessas características.

Análise:

BPM (Beats Per Minute):

- Alto: 122.18 streams
- Baixo: 122.52 streams
- Análise: A diferença é mínima (0.34 streams), indicando que o BPM não tem um impacto significativo na quantidade de streams.



Danceability:

- Alto: 68.22 streams
- Baixo: 66.16 streams
- Análise: Músicas com alta danceability têm uma média ligeiramente maior de streams, mas a diferença (2.06 streams) é pequena.

Energy:

- Alto: 65.14 streams
- Baixo: 63.57 streams
- Análise: Semelhante à danceability, a diferença é mínima (1.57 streams).

Valence:

- Alto: 52.73 streams
- Baixo: 50.06 streams
- Análise: Músicas mais positivas (alta valence) têm uma média um pouco maior de streams, mas a diferença (2.67 streams) ainda é pequena.

Acousticness:

- Alto: 25.69 streams
- Baixo: 27.74 streams
- Análise: Músicas com baixa acousticness (mais eletrônicas) têm uma média ligeiramente maior de streams, mas a diferença (2.05 streams) é pequena.

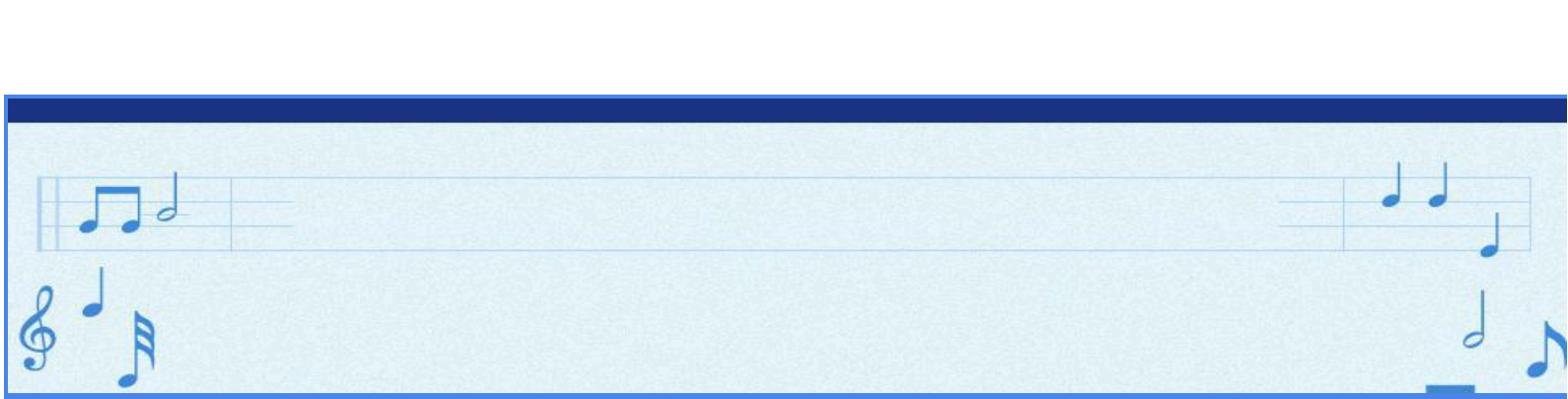
Liveness:

- Alto: 17.41 streams
- Baixo: 19.05 streams
- Análise: Músicas com baixa liveness têm uma média maior de streams, mas a diferença (1.64 streams) é mínima.

Speechiness:

- Alto: 9.06 streams
- Baixo: 11.49 streams
- Análise: Músicas com baixa speechiness (menos palavras faladas) têm uma média maior de streams, mas a diferença (2.43 streams) é pequena.

Instrumentalness:

- 
- Alto: 1.74 streams
 - Baixo: 1.54 streams
 - Análise: A diferença é insignificante (0.20 streams), indicando que a instrumentalidade não afeta significativamente os streams.

Conclusão

As características individuais das músicas (BPM, danceability, energy, valence, acousticness, liveness, speechiness, instrumentalness) não parecem influenciar significativamente a quantidade de streams que elas recebem. A variação na quantidade média de streams entre valores altos e baixos dessas características é mínima, sugerindo que outros fatores podem ser mais importantes para determinar o sucesso de uma música em termos de streams. **Hipótese Refutada!**

Sugestões

Recomendamos que o artista tenha uma estratégia de lançamento e promoção musical baseado nos seguintes fatores :

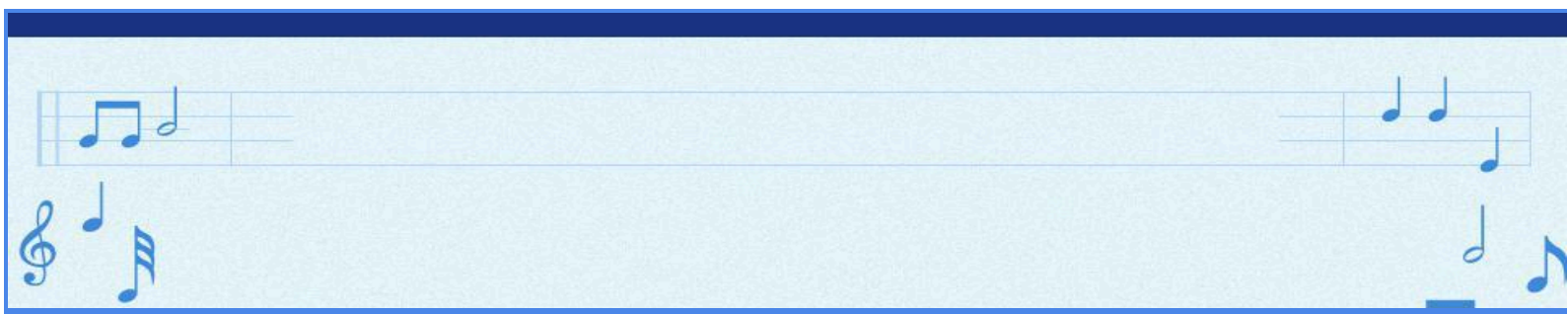
- Procurar especialistas de playlists para promover músicas em playlists populares.
- Realizar lançamentos frequentes para manter uma presença constante no mercado.
- Promover e distribuir músicas em várias plataformas de streaming além do Spotify para alcançar um público maior.
- Focar em qualidade das músicas ao invés de características específicas que não impactam diretamente na popularidade.

Análises Futuras

Existem fatores que podem influenciar o número de streams de um artista, são externos ao Spotify e não podem ser avaliados devido à limitação dos dados disponíveis.

É importante considerar a influência de clipes, desafios, danças virais, TikTok Dances, entre outros, em plataformas de vídeo ou redes sociais, e como isso impacta a quantidade de streams de um artista.

Analisar a música com base em sua data de lançamento e/ou eventos importantes também pode ser útil para estratégias de lançamentos futuros.



Além disso, a coleta de informações adicionais, como gênero musical, país de origem, datas sazonais, entre outros, pode proporcionar uma compreensão mais abrangente e detalhada.

2. Tomada de decisões e informações úteis

- JOIN para unir as três bases de dados.
- Criei a tabela Tab_gravadora

```
# JOIN para unir as tres tabelas spotify, competition e technical_info e criar a Tab_Gravadora.
SELECT
| *
| FROM
| `projeto--2.Hipoteses.Spotify` AS spotify
JOIN
| `projeto--2.Hipoteses.Technical_info` AS technical_info
| ON
| spotify.track_id = technical_info.track_id
JOIN
| `projeto--2.Hipoteses.Competition` AS competition
| ON
| spotify.track_id = competition.track_id;
```

- WITH para tabela temporária.
- Criei uma coluna com ano, mês e dia.
- Substitui o Track_id de valor 0:00 pelo valor mínimo de track_id 1001427 e chamei a nova coluna de Track_id_limpo.
- UPEER para deixar maiúsculos as variáveis string.
- REGEXP_REPLACE para substituir caracteres especiais nas colunas track_name e artist_s__name.
- SAFE_CAST para transformar a coluna streams em número inteiros.
- Somei as colunas de in_spotify_playlists + in_apple_playlists + in_deezer_playlists
- Calculei a média das colunas in_apple_charts, in_deezer_charts, in_shazam_charts e in_spotify_charts.
- Criei a tabela Tab_gravadora1.



```
#Limpeza dos dados da Tab_Gravadora e criaçao da tabela Tab_Gravadora1.
WITH Gravadora AS (
  SELECT
    *,
    DATE(released_year, released_month, released_day) AS released_date,
    CASE
      WHEN track_id = '0:00' THEN '1001427'
      ELSE track_id
    END AS track_id_limpo,
    UPPER(mode) AS mode_maiusculo,
    UPPER(REGEXP_REPLACE(track_name, '[^a-zA-Z0-9 ]', '')) AS track_name_limpo,
    UPPER(REGEXP_REPLACE(artist_s_name, '[^a-zA-Z0-9 ]', '')) AS artist_name_limpo,
    SAFE_CAST(streams AS INT64) AS streams_limpo,
    (in_spotify_playlists + in_apple_playlists + in_deezer_playlists) AS total_playlists,
    (IFNULL(in_apple_charts, 0) + IFNULL(in_deezer_charts, 0) + IFNULL(in_shazam_charts, 0) + IFNULL(in_spotify_charts, 0)) / 4.0 AS media_total_charts
  FROM
    `projeto--2.Hipoteses.Tab_Gravadora`
  WHERE
    in_shazam_charts IS NOT NULL
)
SELECT *
FROM Gravadora;
```

- Alterei o valor nulo presente na coluna Streams para a média de streams no valor de 514137424 e criei uma nova coluna chamada streams_corrigido.
- Criei a tabela Tab_gravadora2.

```
#Alterar o valor nulo da coluna stream_limpo para a media na Tab_Gravadora1 e criaçao da tabela Tab_Gravadora2.
SELECT *,
  CASE
    WHEN streams_limpo IS NULL THEN 514137424
    ELSE streams_limpo
  END AS streams_corrigido
FROM `projeto--2.Hipoteses.Tab_Gravadora1`
```

- Criei uma nova variável de quantidade de músicas por artista.
- Criei a tabela Tab_gravadora3.

```
#WITH para criar tabela temporária para calcular o numero de musica de cada artista da Tab_Gravadora2 e criaçao da tabela Tab_Gravadora3.
WITH n_musicas AS (
  SELECT
    artist_name_limpo,
    COUNT(*) AS n_musicas
  FROM `projeto--2.Hipoteses.Tab_Gravadora2`
  GROUP BY
    artist_name_limpo
)
SELECT *
FROM `projeto--2.Hipoteses.Tab_Gravadora2` AS Gravadora
RIGHT JOIN n_musicas
ON Gravadora.artist_name_limpo = n_musicas.artist_name_limpo;
```

- Utilizei EXCEPT para excluir colunas repetidas.
- Criei a tabela Tab_gravadora_Final.

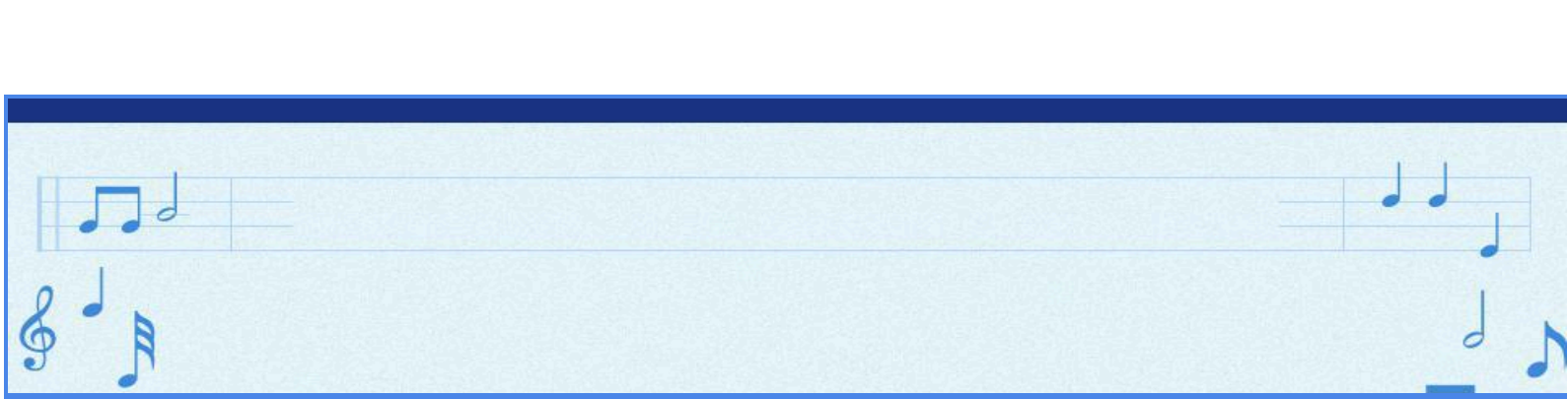
```
#EXCEPT para limpeza das colunas repetidas da Tab_Gravadora3 e criaçao da tabela Tab_Gravadora_Final.
SELECT * EXCEPT (key, track_id_1, track_id_2, track_name, artist_s_name, streams, mode, streams_limpo, artist_name_limpo_1)
FROM `projeto--2.Hipoteses.Tab_Gravadora3`;
```

- Utilizei CREATE OR REPLACE TABLE para substituir a tabela Tab_Gravadora_Final.

- Utilizei a fórmula NTILE para calcular quartis das variáveis categóricas das músicas (streams_corrigido, bpm, danceability__, valence__, energy__, acousticness__, instrumentalness__, liveness__, speechiness__,).
- Segmentar as variáveis categóricas em duas categorias: Alta e Baixa.

```
#Quartil e Segmentação
CREATE OR REPLACE TABLE `projeto--2.Hipoteses.Tab_Gravadora_Final` AS
WITH Quartil AS (
  SELECT
    streams_corrigido, bpm, danceability__, valence__, energy__, acousticness__, instrumentalness__, liveness__, speechiness__,
    NTILE(4) OVER (ORDER BY streams_corrigido) AS quartil_streams,
    NTILE(4) OVER (ORDER BY bpm) AS quartil_bpm,
    NTILE(4) OVER (ORDER BY danceability__) AS quartil_danceability,
    NTILE(4) OVER (ORDER BY valence__) AS quartil_valence,
    NTILE(4) OVER (ORDER BY energy__) AS quartil_energy,
    NTILE(4) OVER (ORDER BY acousticness__) AS quartil_acousticness,
    NTILE(4) OVER (ORDER BY instrumentalness__) AS quartil_instrumentalness,
    NTILE(4) OVER (ORDER BY liveness__) AS quartil_liveness,
    NTILE(4) OVER (ORDER BY speechiness__) AS quartil_speechiness
  FROM `projeto--2.Hipoteses.Tab_Gravadora_Final` AS gravadora
)
SELECT
  gravadora.*,
  Quartil.quartil_streams,
  Quartil.quartil_bpm,
  Quartil.quartil_danceability,
  Quartil.quartil_valence,
  Quartil.quartil_energy,
  Quartil.quartil_acousticness,
  Quartil.quartil_instrumentalness,
  Quartil.quartil_liveness,
  Quartil.quartil_speechiness,
  CASE
    WHEN quartil_streams = 1 OR quartil_streams = 2 THEN 'Baixo'
    WHEN quartil_streams = 3 OR quartil_streams = 4 THEN 'Alto'
  END AS segmentacao_streams,
  CASE
    WHEN quartil_bpm = 1 OR quartil_bpm = 2 THEN 'Baixo'
    WHEN quartil_bpm = 3 OR quartil_bpm = 4 THEN 'Alto'
  END AS segmentacao_bpm,
  CASE
    WHEN quartil_danceability = 1 OR quartil_danceability = 2 THEN 'Baixo'
    WHEN quartil_danceability = 3 OR quartil_danceability = 4 THEN 'Alto'
  END AS segmentacao_danceability,
  CASE
```

- Utilizei a fórmula CORR para calcular a correlação entre os streams e as variáveis categóricas. Hipóteses 1, 2 ,3 e 5.
- Criei uma nova tabela chamada Tab_Gravadora_Correlacao



#Correlação Hipótese 1, 2, 3 e 5.

```
SELECT
CORR (streams_corrigido, bpm) AS corre_streams_bpm,
CORR (in_spotify_charts, in_deezer_charts) AS corre_charts_DeeSpot,
CORR (streams_corrigido, total_playlists) AS corre_streams_playlists,
CORR (streams_corrigido, danceability__) AS corre_streams_dancea,
CORR (streams_corrigido, valence__) AS corre_streams_valen,
CORR (streams_corrigido, energy__) AS corre_streams_energ,
CORR (streams_corrigido, acousticness__) AS corre_streams_acoust,
CORR (streams_corrigido, instrumentalness__) AS corre_streams_instru,
CORR (streams_corrigido, liveness__) AS corre_streams_liven,
CORR (streams_corrigido, speechiness__) AS corre_streams_speech
FROM `projeto--2.Hipoteses.Tab_Gravadora_Final`;
```

- Cálculo de quantidade de músicas por artista.
- Criei nova tabela chamada Tab_Gravadora_N_Musicas.

verificar a quantidade de músicas - Tabela TAB_Gravadora_N_Musica

```
SELECT
  artist_name_limpo,
  COUNT(DISTINCT track_name_limpo) AS total_musicas,
  SUM(streams_corrigido) AS total_streams
FROM
  `projeto--2.Hipoteses.Tab_Gravadora_Final`
GROUP BY
  artist_name_limpo
```

- Utilizei a fórmula CORR para calcular a correlação entre os streams e número de músicas por artista. Hipótese 4.

Ferramentas utilizadas

- BigQuery
- Power BI
- Google Colab
- Google Apresentações

Linguagens utilizadas

- SQL
- Python
- DAX



Links úteis

- [BigQuery](#)
- [PowerBI](#)
- [Apresentação](#)
- [Base de Dados](#)
- [Vídeo de Apresentação](#)