



## **Análise das Músicas Mais Reproduzidas no Spotify (2024)**

**Daniel dos Santos da Silva, Enzo Ferroni, Hudson Carvalho de Jesus, Vinícius de Souza Sabiá**

{10720767@mackenzista.com.br, 10417100@mackenzista.com.br,  
10721459@mackenzista.com.br, 10721475@mackenzista.com.br}

<https://github.com/EnzoFerroni/ProjetoAplicado1>



## Sumário

1. Introdução .....	3
2. Glossário .....	3
3. Contexto do Estudo .....	4
4. Objetivo do Estudo .....	4
5. Apresentação da empresa e problema de pesquisa .....	4
6. Referências do Dataset .....	6
7. Descrição do Dataset .....	6
8. Apresentação dos Metadados .....	6
9. Análise Exploratória de dados.....	9
10. Conclusão e Próximos Passos.....	14
11. Referências .....	14

### Lista de Tabelas

Tabela 1 – Glossário de termos utilizados na análise de dados

Tabela 2 – Apresentação dos metadados do dataset

### Lista de Imagens

Imagem 1 - Exibição das Primeiras Linhas do Dataset: Tabela com os primeiros registros do dataset.

Imagem 2 - Informações Gerais do Dataset: Exibe estrutura e quantidade de dados ausentes no dataset.

Imagem 3 - Estatísticas Descritivas: Tabela com estatísticas descritivas das variáveis numéricas.

Imagem 4 - Verificação de Valores Nulos: Código e saída mostrando a quantidade de valores ausentes.

Imagem 5 - Remoção de Duplicatas e Análise de Outliers: Código para remover duplicatas e gerar boxplots.

Imagem 6 - Boxplot das Variáveis Numéricas: Gráfico boxplot identificando possíveis outliers nos dados.

Imagem 7 - Distribuição das Variáveis Numéricas: Histogramas das variáveis numéricas do dataset.

Imagem 8 - Matriz de Correlação: Mapa de calor com a matriz de correlação entre as variáveis numéricas.

Imagem 9 - Top 10 Artistas Mais Presentes: Gráfico de barras mostrando os artistas mais frequentes no dataset.

Imagem 10 - Relação entre Popularidade e Streams do Spotify: Gráfico de dispersão relacionando a quantidade de streams com a popularidade no Spotify.



## 1. Introdução

Nos últimos anos, o streaming revolucionou a forma como consumimos música, e o Spotify se tornou uma das principais plataformas nesse cenário. A possibilidade de acessar milhões de faixas a qualquer momento fez com que os hábitos dos ouvintes mudassem drasticamente, criando tendências e redefinindo o conceito de sucesso musical. Este projeto tem como objetivo analisar os dados das músicas mais reproduzidas no Spotify em 2024 para entender quais fatores influenciam seu desempenho. A partir do dataset disponível no Kaggle, faremos uma exploração detalhada para identificar os gêneros musicais mais populares, a relação entre as músicas mais escutadas, os artistas e gravadoras que mais aparecem na lista e etc.

Com essa análise, esperamos obter insights relevantes sobre o comportamento dos ouvintes e as tendências da indústria musical no contexto atual.

## 2. Glossário

Seção técnica para padronizar termos

Termo	Definição
<b>Streams</b>	Quantidade total de vezes que uma música foi reproduzida na plataforma Spotify.
<b>Outlier</b>	Valor que se desvia significativamente da maioria dos dados em um conjunto. Pode indicar erro ou um comportamento atípico relevante.
<b>Missing Value (NA)</b>	Valor ausente ou não informado em determinado campo do dataset. Deve ser tratado para garantir consistência nas análises estatísticas.
<b>Deep Learning</b>	Subcampo do aprendizado de máquina que utiliza redes neurais profundas para processar e identificar padrões complexos em dados como áudio, imagens e texto.
<b>Reinforcement Learning</b>	Técnica de aprendizado de máquina baseada em tentativa e erro, onde um algoritmo aprende a tomar decisões otimizadas ao longo do tempo.
<b>Testes A/B</b>	Experimentos estatísticos realizados para comparar diferentes versões de um modelo ou funcionalidade e determinar qual apresenta melhor desempenho.
<b>Filtragem Colaborativa</b>	Método de recomendação baseado no comportamento e preferências de usuários semelhantes para sugerir conteúdo relevante.
<b>TensorFlow</b>	Framework de código aberto amplamente utilizado para construir e treinar modelos de aprendizado de máquina e deep learning.
<b>PySpark</b>	Biblioteca para processamento distribuído de grandes volumes de dados, amplamente usada em análise de big data e machine learning.



Termo	Definição
<b>Personalização Dinâmica</b>	Processo de ajuste contínuo das recomendações com base em interações e preferências em tempo real do usuário.

### 3. Contexto do Estudo

A indústria da música tem passado por transformações intensas nos últimos anos, e as plataformas de streaming desempenham um papel central nesse processo. Diferente da era dos CDs ou do download digital, hoje o sucesso de uma música é medido principalmente pelo número de reproduções que ela recebe.

O Spotify, sendo uma das plataformas mais influentes, pode revelar muito sobre as preferências do público global. Por isso, analisar os dados das músicas mais reproduzidas em 2024 nos permite entender melhor os fatores que contribuem para o sucesso de um artista ou faixa.

Essa análise pode ser útil não apenas para pesquisadores e profissionais de dados, mas também para artistas, produtores e gravadoras que buscam insights sobre o mercado e as preferências do público.

### 4. Objetivo do Estudo

Com base no dataset *Most\_Streamed\_Spotify\_Songs\_2024.csv*, o objetivo é identificar os fatores que impulsionam o sucesso de músicas no Spotify em 2024, utilizando métricas como:

- Spotify Streams (principal indicador de popularidade),
- Spotify Playlist Reach (alcance em playlists),
- Spotify Playlist Count (número de playlists que incluem a música),
- Spotify Popularity (pontuação de popularidade da plataforma),
- YouTube Views e TikTok Views (engajamento em outras plataformas),

Metodologia:

1. Análise Descritiva:
  - Ranking das músicas mais streamed e correlação entre variáveis (ex: streams × playlists).
  - Comparação de médias (ex: músicas explícitas vs. não explícitas).
2. Análise Preditiva (se aplicável):
  - Identificar variáveis com maior impacto nos streams (ex: playlists têm mais peso que YouTube?).
3. Visualização:
  - Gráficos de dispersão (streams × playlist reach), heatmaps de correlação e rankings.

### 5. Apresentação da empresa e problema de pesquisa

O Spotify é uma das maiores plataformas de streaming de áudio do mundo, fundada em 2006 na Suécia com a missão de conectar artistas e ouvintes através de uma experiência musical personalizada. A empresa opera com um modelo de negócios freemium, oferecendo um serviço gratuito suportado por anúncios e um serviço premium baseado em assinatura. Atualmente, conta com mais de 574 milhões de usuários ativos e detém



31% de market share global, superando concorrentes como Apple Music, Amazon Music e YouTube Music.

O Spotify se destaca por sua abordagem data-driven, utilizando algoritmos de recomendação avançados como o Discover Weekly e o Spotify Wrapped para engajar seu público. Seu diferencial competitivo está na personalização da experiência do usuário, utilizando machine learning e inteligência artificial para identificar padrões de comportamento e preferências musicais. Seus valores de inovação, diversidade e paixão pela música guiam iniciativas como o Spotify for Artists, que fornece insights detalhados para artistas, e parcerias estratégicas, incluindo colaborações com IA generativa, como o DJ AI, que aprimora a descoberta de conteúdo e a retenção de usuários.

### **Market Share e Posição no Mercado**

Líder no setor de streaming de áudio, o Spotify possui cerca de 236 milhões de assinantes premium e 602 milhões de usuários ativos globalmente (dados de 2024). Seu alcance global se estende por mais de 180 países, consolidando sua posição como a principal plataforma do segmento.

### **Desafios e Oportunidades**

O principal desafio do Spotify é melhorar a personalização para reduzir a taxa de cancelamento. Para isso, a empresa investe fortemente em análise de dados e machine learning, buscando entender padrões de escuta e preferências musicais dos usuários. Projetos como análise de sentimento em letras e modelos preditivos de sucesso musical reforçam seu compromisso com a inovação e ajudam na curadoria automatizada de playlists personalizadas.

Além disso, a análise dos dados do Spotify revela padrões consistentes no comportamento dos usuários, como preferências musicais que variam conforme o horário do dia, dia da semana e até mesmo estado emocional. Por exemplo, há um aumento do consumo de músicas animadas durante as manhãs e finais de semana, enquanto canções mais calmas e introspectivas são mais ouvidas à noite. Eventos culturais e lançamentos de artistas também geram picos específicos de consumo, demonstrando que há uma forte correlação entre contexto externo e hábitos de escuta.

O Spotify já utiliza esses insights em seus algoritmos de recomendação, como o Discover Weekly, que combina filtragem colaborativa e análise de conteúdo para personalizar playlists. Porém, novas abordagens podem otimizar ainda mais a experiência do usuário e aumentar a retenção.

### **Iniciativas em Data Science**

O Spotify investe continuamente em tecnologias avançadas de análise de dados para aprimorar a experiência do usuário, aplicando modelos de Deep Learning para análise de áudio e letras, além de utilizar modelos preditivos para identificar tendências musicais e prever o sucesso de novas faixas antes mesmo de seu lançamento. A plataforma também incorpora Reinforcement Learning, permitindo que as recomendações sejam ajustadas dinamicamente em tempo real para garantir maior personalização. Para validar e otimizar esses modelos, o Spotify realiza extensivos testes A/B, assegurando que as sugestões



musicais sejam cada vez mais precisas e relevantes para cada usuário.

O desenvolvimento contínuo de novos algoritmos torna o sistema mais eficiente, permitindo que o Spotify não apenas antecipe tendências musicais, mas também refine a personalização das recomendações de forma mais estratégica. Tecnologias como TensorFlow e PySpark possibilitam o processamento de grandes volumes de dados, enquanto modelos híbridos que integram a análise de áudio, letras e comportamento do usuário ampliam significativamente a capacidade da plataforma de oferecer uma experiência única e adaptável.

Com esses investimentos, o Spotify transforma dados em vantagem competitiva, consolidando sua posição como líder no setor de streaming e revolucionando continuamente a forma como o mundo descobre e consome música.

## 6. Referências do Dataset

Para realizar este estudo, utilizaremos o dataset "Most Streamed Spotify Songs 2024", disponível no Kaggle.

**Fonte:** [Kaggle - Most Streamed Spotify Songs 2024](#)

**Restrições de Uso:** Os dados são de acesso público e podem ser utilizados para fins educacionais e acadêmicos.

**Período da Coleta:** 2024

## 7. Descrição do Dataset

O dataset "Most Streamed Spotify Songs 2024" foi criado e disponibilizado por Nel Giri Yewithana na plataforma Kaggle. Ele foi construído a partir de dados extraídos do Spotify, utilizando a API oficial da plataforma para coletar informações sobre as músicas mais reproduzidas no ano de 2024.

O objetivo da criação desse dataset foi fornecer um recurso abrangente para pesquisadores, analistas de dados e entusiastas da música, permitindo a exploração de padrões musicais, tendências de streaming e características sonoras das faixas mais populares.

Como os dados foram extraídos diretamente da API do Spotify, eles refletem informações oficiais da plataforma, incluindo métricas de popularidade, características acústicas das músicas e estatísticas de reprodução. Dessa forma, o dataset serve como uma fonte confiável para análises sobre o comportamento dos ouvintes e o desempenho das faixas no mercado musical global.

## 8. Apresentação dos Metadados



Atributos	Definição	Relevância para o estudo	Exemplo no Dataset
Track	Nome da música (String, sem limites).	Identifica a música no ranking.	MILLION DOLLAR BABY
Album Name	Nome do álbum (String, sem limites).	Relaciona a música ao seu álbum de origem.	Million Dollar Baby - Single
Artist	Nome do artista (String, sem limites).	Permite análise de popularidade por artista.	Tommy Richman
Release Date	Data de lançamento (String, formato DD/MM/AAAA).	Ajuda a identificar tendências ao longo do tempo.	26/04/2024
ISRC	Código único da música (String, sem limites).	Importante para rastrear a música em diferentes plataformas.	QM24S2402528
All Time Rank	Posição no ranking global (Inteiro, 1 - 4600).	Indica a popularidade da música.	1
Track Score	Pontuação da música (Float, 0 - 725.4).	Mede a influência da música no cenário musical.	725,4
Spotify Streams	Número de reproduções no Spotify (Inteiro, 0 - 2,031,280,633).	Indica o sucesso da música na plataforma mais usada de streaming.	390470936
Spotify Playlist Count	Número de playlists no Spotify (Inteiro, 0 - 269,802).	Mede a aceitação da música em curadorias.	30716
Spotify Playlist Reach	Alcance total das playlists no Spotify (Inteiro, 0 - 211,607,669).	Estima a exposição da música dentro da plataforma.	196631588



Spotify Popularity	Índice de popularidade no Spotify (Float, 0 - 100).	Avalia a relevância da música na plataforma.	92
YouTube Views	Visualizações no YouTube (Inteiro, 0 - 1,096,100,899).	Mede o impacto da música na plataforma de vídeo.	84274754
YouTube Likes	Curtidas no YouTube (Inteiro, 0 - 10,629,796).	Mede o engajamento na plataforma.	651565900
TikTok Views	Visualizações no TikTok (Inteiro, 0 - 14,603,725,994).	Representa a exposição da música na rede.	5332281936
YouTube Playlist Reach	Alcance das playlists no YouTube (Inteiro, 0 - 3,351,188,582).	Mede a disseminação via curadoria.	150597040
Apple Music Playlist Count	Número de playlists na Apple Music (Float, 0 - 394).	Avalia a aceitação da música entre usuários da Apple.	210
AirPlay Spins	Execuções em rádios (Inteiro, 0 - 1,474,799).	Mede a presença da música nas rádios.	40975
SiriusXM Spins	Execuções na SiriusXM (Inteiro, 0 - 2,182).	Mede a popularidade em rádios premium.	684
Deezer Playlist Count	Número de playlists na Deezer (Float, 0 - 264).	Mede a inclusão na plataforma.	62
Deezer Playlist Reach	Alcance das playlists da Deezer (Inteiro, 0 - 36,321,847).	Mede a disseminação na plataforma.	17598718
Amazon Playlist Count	Número de playlists na Amazon Music (Float, 0 - 210).	Mede a aceitação na Amazon Music.	114





Pandora Streams	Reproduções no Pandora (Inteiro, 0 - 190,260,277).	Mostra a presença da música na plataforma.	18004655
Pandora Track Stations	Estações do Pandora com a música (Inteiro, 0 - 203,384).	Indica a curadoria da música.	22931
Soundcloud Streams	Reproduções no SoundCloud (Inteiro, 0 - 7,208,651).	Mede a popularidade na plataforma.	4818457
Shazam Counts	Identificações no Shazam (Inteiro, 0 - 11,822,942).	Indica o interesse do público em descobrir a música.	2669262
TIDAL Popularity	Popularidade da música no TIDAL (Float, 0 - 100).	Mede o impacto da música entre usuários do TIDAL.	N/A
Explicit Track	Indica se a música tem conteúdo explícito (Inteiro, 0 = Não, 1 = Sim).	Importante para análise de restrições e públicos-alvo.	0

## 9. Análise Exploratória de dados

Esta seção apresenta os métodos e procedimentos utilizados na análise exploratória de dados. O conteúdo será exposto em um Google Colab Notebook, integrando código e comentários explicativos para facilitar a compreensão do tema.

	Track	Album Name	Artist	Release Date	ISRC	All Time Rank	Track Score	Spotify Streams	Spotify Playlist Count	Spotify Playlist Reach	...	SiriusXM Spins	Deezer Playlist Count	DPI
0	MILLION DOLLAR BABY	Million Dollar Baby - Single	Tommy Richman	4/26/2024	QM24S2402528	1	725.4	390,470,936	30,716	196,631,588	...	684	62.0	17,55
1	Not Like Us	Not Like Us	Kendrick Lamar	5/4/2024	USUG12400910	2	545.9	323,703,884	28,113	174,597,137	...	3	67.0	10,42
2	i like the way you kiss me	I like the way you kiss me	Artemas	3/19/2024	QZJ842400387	3	538.4	601,309,283	54,331	211,607,669	...	536	136.0	36,32
3	Flowers	Flowers - Single	Miley Cyrus	1/12/2023	USSM12209777	4	444.9	2,031,280,633	269,802	136,569,078	...	2,182	264.0	24,66
4	Houdini	Houdini	Eminem	5/31/2024	USUG12403398	5	423.3	107,034,922	7,223	151,469,874	...	1	82.0	17,66



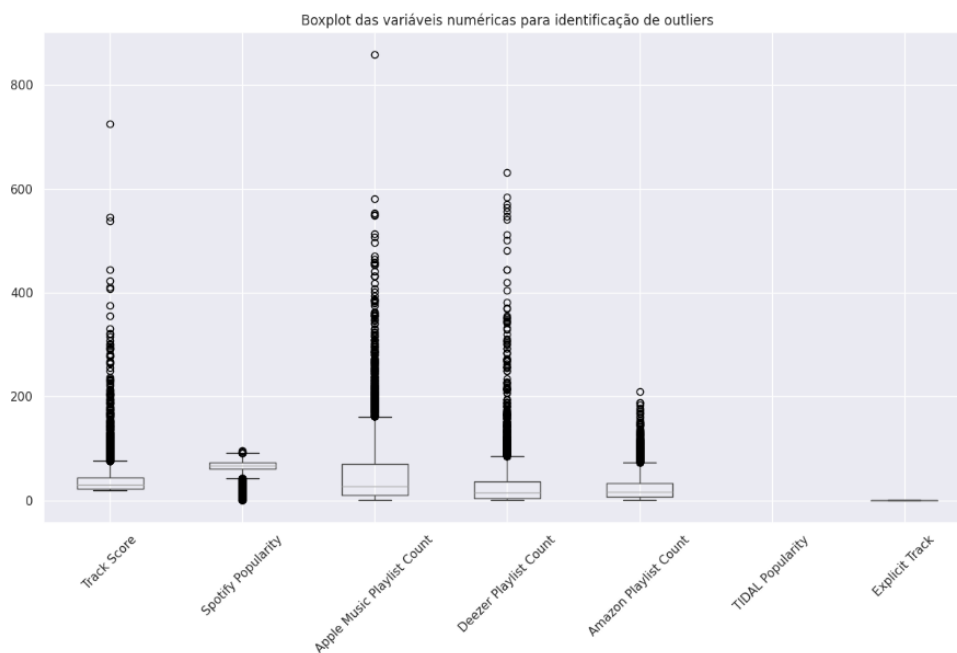
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4600 entries, 0 to 4599
Data columns (total 29 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Track                                4600 non-null   object
1   Album Name                           4600 non-null   object
2   Artist                               4595 non-null   object
3   Release Date                         4600 non-null   object
4   ISRC                                 4600 non-null   object
5   All Time Rank                       4600 non-null   object
6   Track Score                          4600 non-null   float64
7   Spotify Streams                      4487 non-null   object
8   Spotify Playlist Count               4530 non-null   object
9   Spotify Playlist Reach               4528 non-null   object
10  Spotify Popularity                   3796 non-null   float64
11  YouTube Views                       4292 non-null   object
12  YouTube Likes                       4285 non-null   object
13  TikTok Posts                        3427 non-null   object
14  TikTok Likes                        3620 non-null   object
15  TikTok Views                        3619 non-null   object
16  YouTube Playlist Reach               3591 non-null   object
17  Apple Music Playlist Count           4039 non-null   float64
18  AirPlay Spins                       4102 non-null   object
19  SiriusXM Spins                      2477 non-null   object
20  Deezer Playlist Count                3679 non-null   float64
21  Deezer Playlist Reach                3672 non-null   object
22  Amazon Playlist Count                3545 non-null   float64
23  Pandora Streams                     3494 non-null   object
24  Pandora Track Stations               3332 non-null   object
25  Soundcloud Streams                  1267 non-null   object
26  Shazam Counts                       4023 non-null   object
27  TIDAL Popularity                     0 non-null      float64
28  Explicit Track                       4600 non-null   int64
dtypes: float64(6), int64(1), object(22)
memory usage: 1.0+ MB
```

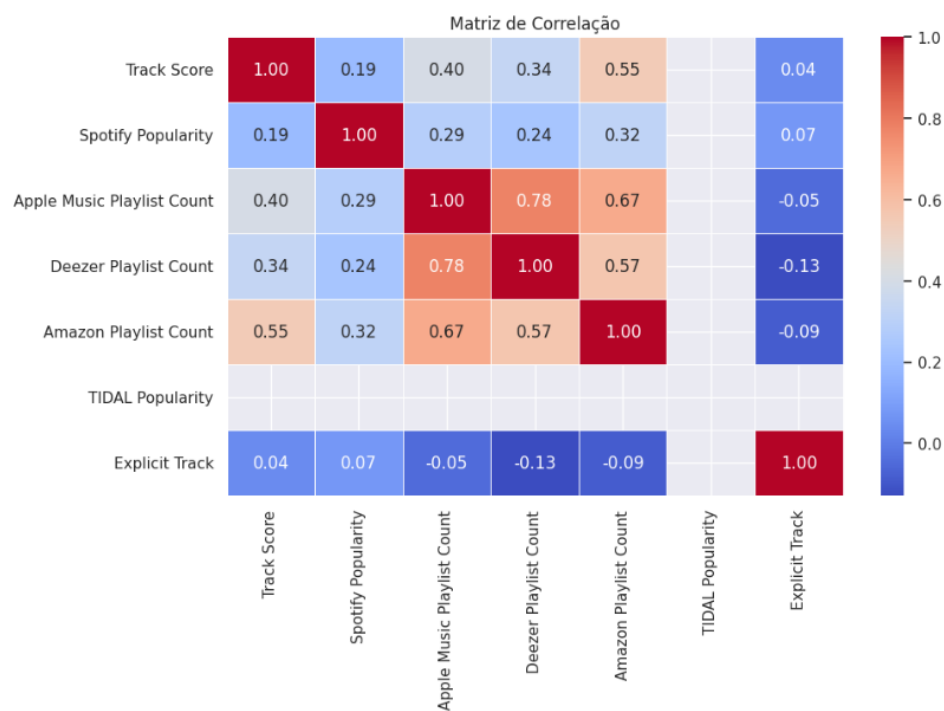
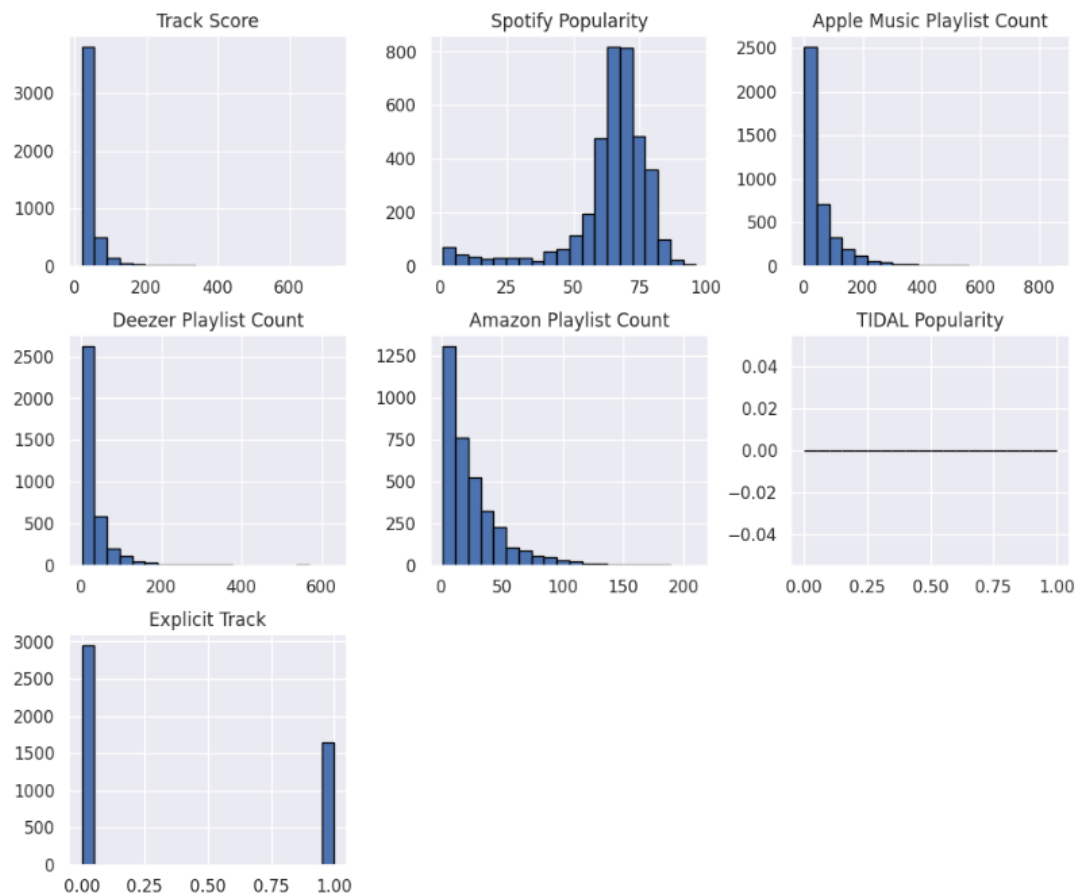
	Track Score	Spotify Popularity	Apple Music Playlist Count	Deezer Playlist Count	Amazon Playlist Count	TIDAL Popularity	Explicit Track
count	4600.000000	3796.000000	4039.000000	3679.000000	3545.000000	0.0	4600.000000
mean	41.844043	63.501581	54.60312	32.310954	25.348942	NaN	0.358913
std	38.543766	16.186438	71.61227	54.274538	25.989826	NaN	0.479734
min	19.400000	1.000000	1.00000	1.000000	1.000000	NaN	0.000000
25%	23.300000	61.000000	10.00000	5.000000	8.000000	NaN	0.000000
50%	29.900000	67.000000	28.00000	15.000000	17.000000	NaN	0.000000
75%	44.425000	73.000000	70.00000	37.000000	34.000000	NaN	1.000000
max	725.400000	96.000000	859.00000	632.000000	210.000000	NaN	1.000000

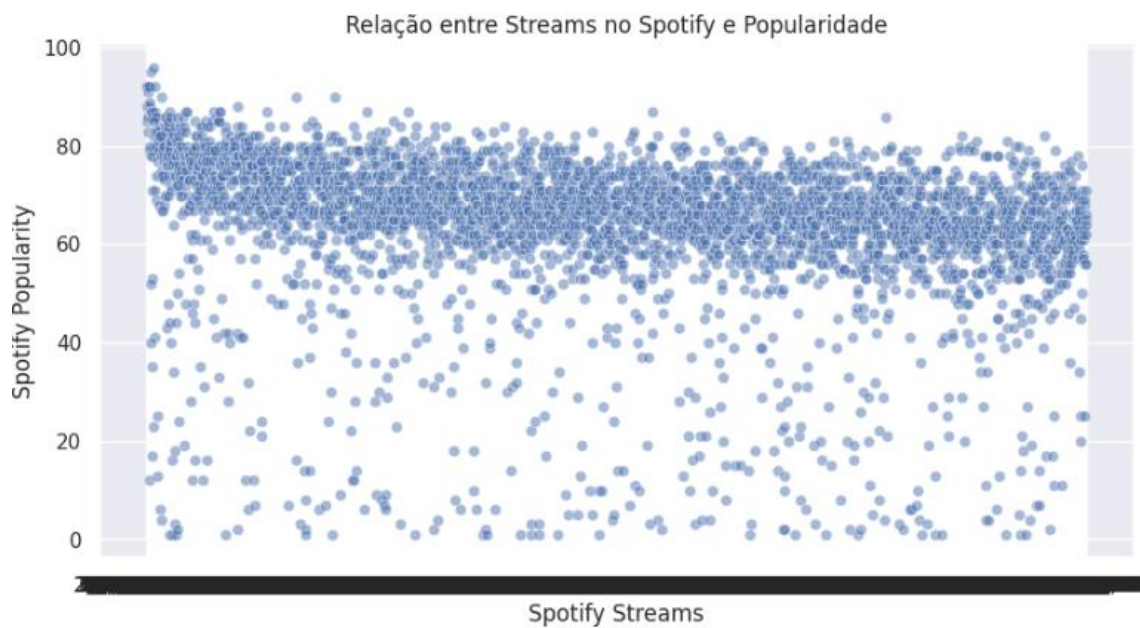
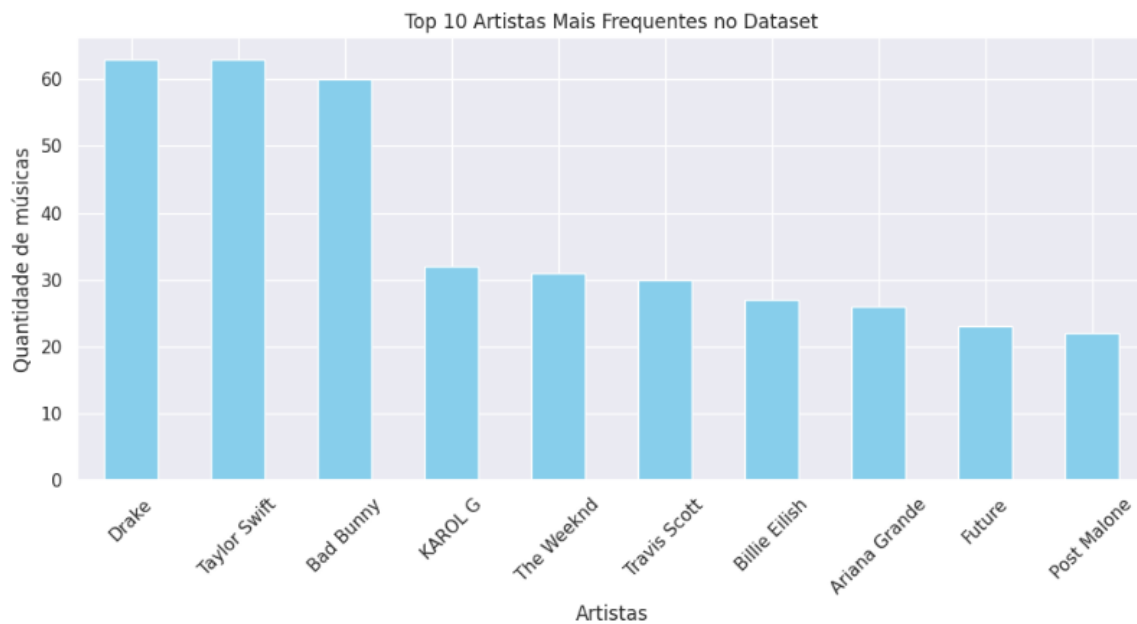


```
Track 0
Album Name 0
Artist 5
Release Date 0
ISRC 0
All Time Rank 0
Track Score 0
Spotify Streams 113
Spotify Playlist Count 70
Spotify Playlist Reach 72
Spotify Popularity 804
YouTube Views 308
YouTube Likes 315
TikTok Posts 1173
TikTok Likes 980
TikTok Views 981
YouTube Playlist Reach 1009
Apple Music Playlist Count 561
AirPlay Spins 498
SiriusXM Spins 2123
Deezer Playlist Count 921
Deezer Playlist Reach 928
Amazon Playlist Count 1055
Pandora Streams 1106
Pandora Track Stations 1268
Soundcloud Streams 3333
Shazam Counts 577
TIDAL Popularity 4600
Explicit Track 0
dtype: int64
```

Valores duplicados no dataset: 2







O eixo X do gráfico que relaciona Streams e Popularidade no Spotify indica que, quanto mais à direita um ponto estiver, maior é o número de streams da música.



## 10. Conclusão e Próximos Passos

A análise dos dados das músicas mais reproduzidas no Spotify em 2024 revelou padrões importantes sobre as tendências e preferências do público na era do streaming. Utilizando o dataset disponível no Kaggle, foi possível identificar os fatores que influenciam o sucesso de uma faixa, como os gêneros musicais mais populares, os artistas mais influentes e a relevância das gravadoras na promoção de novos sucessos. Esses insights fornecem uma visão detalhada do comportamento do consumidor e representam um valioso recurso para a indústria musical, auxiliando na tomada de decisões estratégicas baseadas em dados.

Além disso, os resultados desta pesquisa contribuem para um entendimento mais amplo sobre como o streaming tem moldado o consumo de música e redefinido o conceito de sucesso na indústria fonográfica. A forma como os usuários interagem com as plataformas digitais, suas preferências em diferentes momentos do dia e o impacto das recomendações algorítmicas são aspectos que se mostram cada vez mais determinantes para a popularidade de uma faixa.

Diante dessas descobertas, um aprofundamento da análise poderia explorar a variação das preferências musicais ao longo do tempo, observando tendências sazonais e o impacto de eventos culturais ou lançamentos estratégicos. Além disso, entender de que maneira a personalização do conteúdo influencia o engajamento do usuário pode trazer novas perspectivas sobre a eficácia das ferramentas de recomendação do Spotify. O estudo também poderia avançar no desenvolvimento de modelos preditivos capazes de estimar o potencial de sucesso de uma música antes do seu lançamento, considerando características sonoras, colaborações entre artistas e padrões de consumo anteriores. Outra possibilidade interessante seria comparar esses dados com outras plataformas de streaming para identificar diferenças no comportamento dos usuários e entender como cada serviço influencia a descoberta e o consumo de música.

Com esses desdobramentos, a pesquisa pode oferecer ainda mais subsídios para compreender a dinâmica do mercado musical, reforçando a importância da análise de dados para acompanhar a constante evolução do setor.

## 11. Referências

PYTHON SOFTWARE FOUNDATION. Pandas Documentation. Disponível em: <https://pandas.pydata.org/docs/>. Acesso em: 04 maio 2025.

WASKOM, Michael L. Seaborn: Statistical Data Visualization. Disponível em: <https://seaborn.pydata.org/>. Acesso em: 04 maio 2025.

HUNTER, John D. Matplotlib: Visualization with Python. Disponível em: <https://matplotlib.org/stable/index.html>. Acesso em: 04 maio 2025.