



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/ProjetoAplicadoI>



Sumário

1. Introdução.....	3
2. Glossário.....	3
3. Contexto do Estudo.....	3
4. Objetivo do Estudo.....	4
5. Apresentação da empresa e problema de pesquisa	4
6. Referências do Dataset.....	5
7. Descrição do Dataset.....	5
8. Apresentação dos Metadados.....	6
9. Análise Exploratória de dados	8
10. Conclusão.....	17



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
Streams:	Número de vezes que uma música foi reproduzida no Spotify;
Track Score:	Nota dada a música conforme vários fatores (19,4-725);
Playlist Count:	Número de playlists que a música está;
Playlist Reach:	Número de playlist;
All time rank:	Ranque de todos os tempos das músicas;
Popularity:	Nota dada a música conforme o spotify (1-96).

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. Com mais de 574 milhões de usuários ativos e 31% de market share global, a empresa se destaca por sua abordagem data-driven, utilizando algoritmos de recomendação como o Discover Weekly e o Spotify Wrapped para engajar seu público. Seus valores de inovação, diversidade e paixão pela música guiam iniciativas como o Spotify for Artists e parcerias com IA generativa, como o DJ AI, que aprimoram a descoberta de conteúdo e a retenção de usuários.

O principal desafio da Spotify é melhorar a personalização para reduzir a taxa de cancelamento, utilizando análise de dados e machine learning para entender padrões de escuta e preferências musicais. Projetos como a análise de sentimento em letras e modelos preditivos de sucesso musical reforçam seu compromisso com a inovação. Como líder do setor, a empresa continua investindo em tecnologias avançadas para manter sua vantagem competitiva, provando que os dados são essenciais para transformar a maneira como o mundo consome música.

A análise dos dados do Spotify revela padrões consistentes no comportamento dos usuários,



como preferências musicais por horário, dia da semana e até mesmo estado emocional. Esses padrões seguem tendências previsíveis, como o aumento do consumo de músicas animadas durante as manhãs e finais de semana, ou a preferência por canções calmas à noite. Além disso, eventos culturais e lançamentos de artistas geram picos de consumo específicos, 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.

O desenvolvimento de novos algoritmos pode otimizar ainda mais a experiência do usuário e aumentar a retenção. Uma abordagem eficaz seria implementar modelos híbridos que integrem técnicas de deep learning para análise de áudio e letras com reinforcement learning para ajustar recomendações em tempo real. Ferramentas como TensorFlow e PySpark permitiriam processar grandes volumes de dados e identificar padrões complexos, enquanto testes A/B validariam a eficácia das melhorias. Com isso, o Spotify não apenas entenderia melhor os hábitos dos usuários, mas também anteciparia suas preferências, transformando dados em vantagem competitiva e fortalecendo sua posição no mercado de streaming.

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.


```
In [1]: # Importando bibliotecas necessárias
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: # Configuração do estilo dos gráficos
sns.set_theme(style="darkgrid")
```

```
In [6]: # Carregando o dataset
file_path = "/content/Most_Streamed_Spotify_Songs_2024.csv"
df = pd.read_csv(file_path, encoding='latin1')
```

Foi necessário usar o 'encoding' pois estava com o erro "UnicodeDecodeError: 'utf-8' codec can't decode byte 0xfd in position 2679: invalid start byte"

```
In [8]: # Primeiras Linhas do dataset
display(df.head())
```

	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	Deezer Playlist Reach	Amazon Playlist Count	Pandora Streams	Pandora Track Stations	Soundcloud Streams	Shazam Counts	TIDAL Popularity	Explicit Track
0	BABY MILLION DOLLAR	Baby - Million Dollar	Richman Tommy	4/26/2024	QM24S2402528	1	725.4	390,470,936	30,716	196,631,588	...	684	62.0	17,598,718	114.0	18,004,655	22,931	4,818,457	2,669,262	NaN	0
1	Not Like	Not	Kendrick	5/4/2024	USUG12400910	2	545.9	323,703,884	28,113	174,597,137	...	3	67.0	10,422,430	111.0	7,780,028	28,444	6,623,075	1,118,279	NaN	1
2	i like the way you kiss me	I like you kiss the way	Artemas	3/19/2024	QZJ842400387	3	538.4	601,309,283	54,331	211,607,669	...	536	136.0	36,321,847	172.0	5,022,621	5,639	7,208,651	5,285,340	NaN	0
3	Flowers	Flowers	Miley	1/12/2023	USSM12209777	4	444.9	2,031,280,633	269,802	136,569,078	...	2,182	264.0	24,684,248	210.0	190,260,277	203,384	NaN	11,822,942	NaN	0
4	Houdini	Houdini	Eminem	5/31/2024	USUG12403398	5	423.3	107,034,922	7,223	151,469,874	...	1	82.0	17,660,624	105.0	4,493,884	7,006	207,179	457,017	NaN	1

5 rows × 29 columns

```
In [9]: # Informações gerais do dataset
df.info()
```

```
<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
```

```
In [10]: # Estadísticas descriptivas
display(df.describe())
```

	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

```
In [11]: # Verificar valores nulos
print(df.isnull().sum())
```

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	

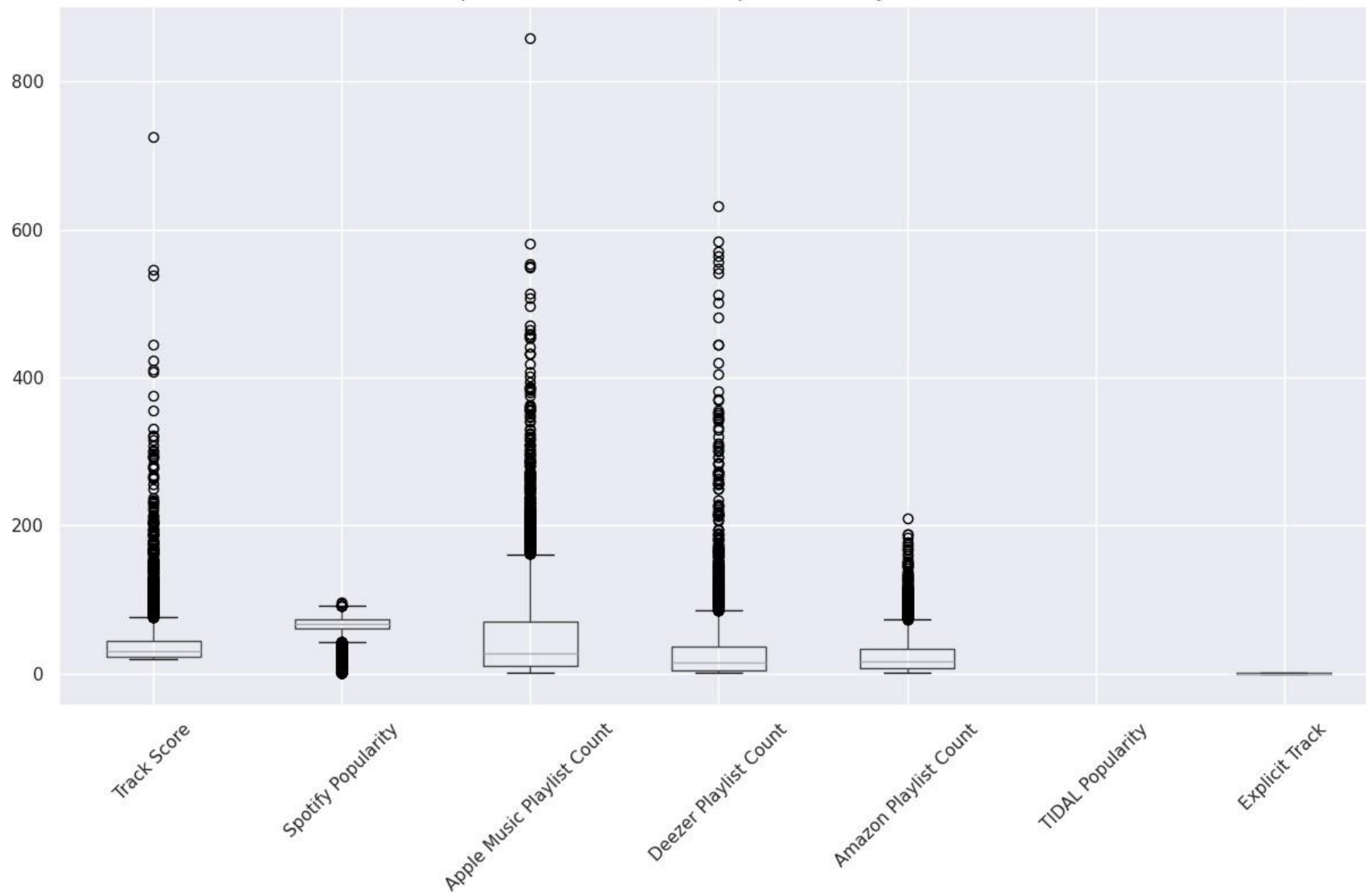
```
In [13]: # Remover duplicatas, se existirem
print("\n Valores duplicados no dataset:", df.duplicated().sum())
df = df.drop_duplicates()
```

Valores duplicados no dataset: 2

```
In [15]: # Análise de Outliers usando Boxplots
numerical_cols = df.select_dtypes(include=['float64', 'int64']).columns

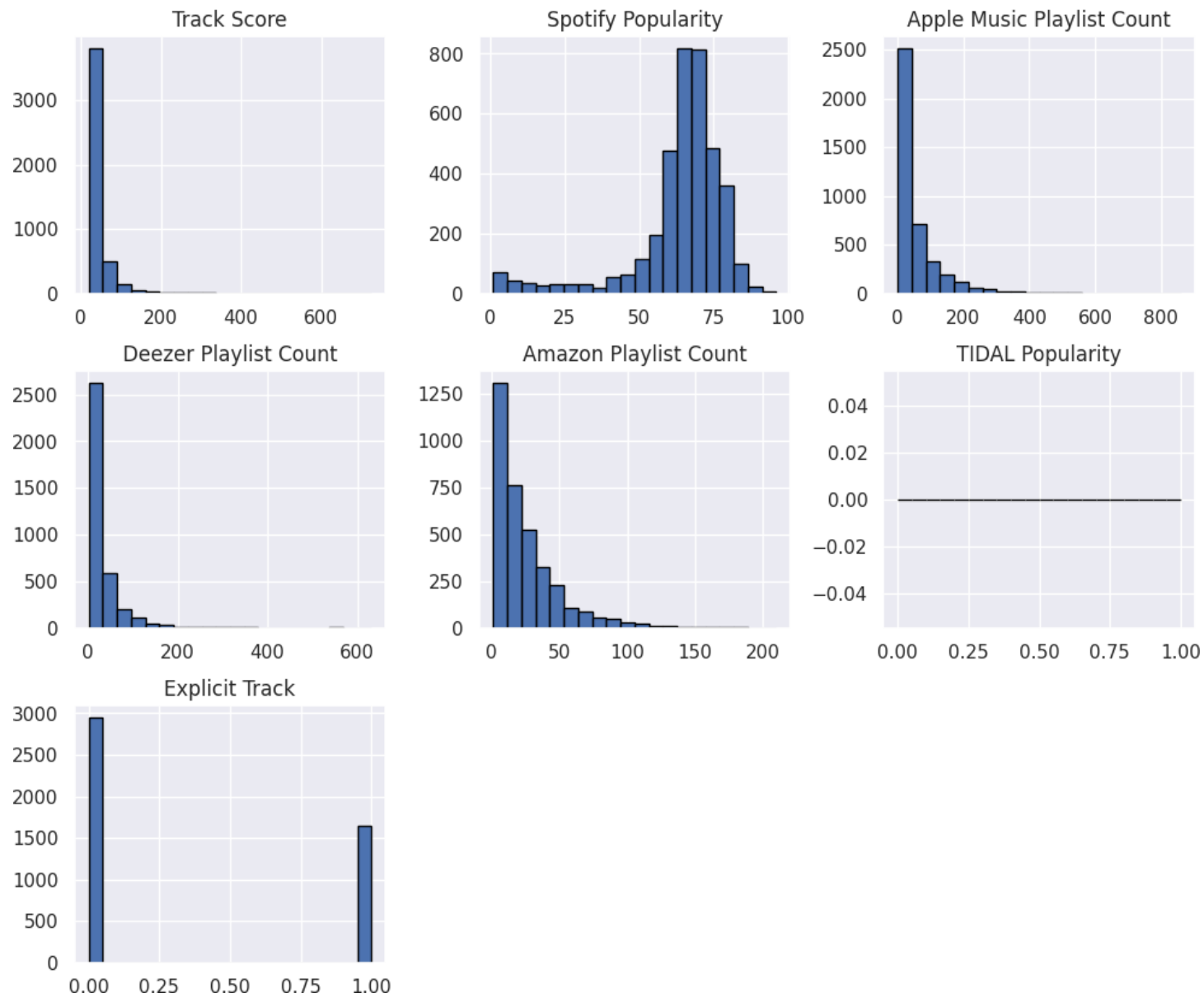
plt.figure(figsize=(15, 8))
df[numerical_cols].boxplot()
plt.xticks(rotation=45)
plt.title("Boxplot das variáveis numéricas para identificação de outliers")
plt.show()
```

Boxplot das variáveis numéricas para identificação de outliers

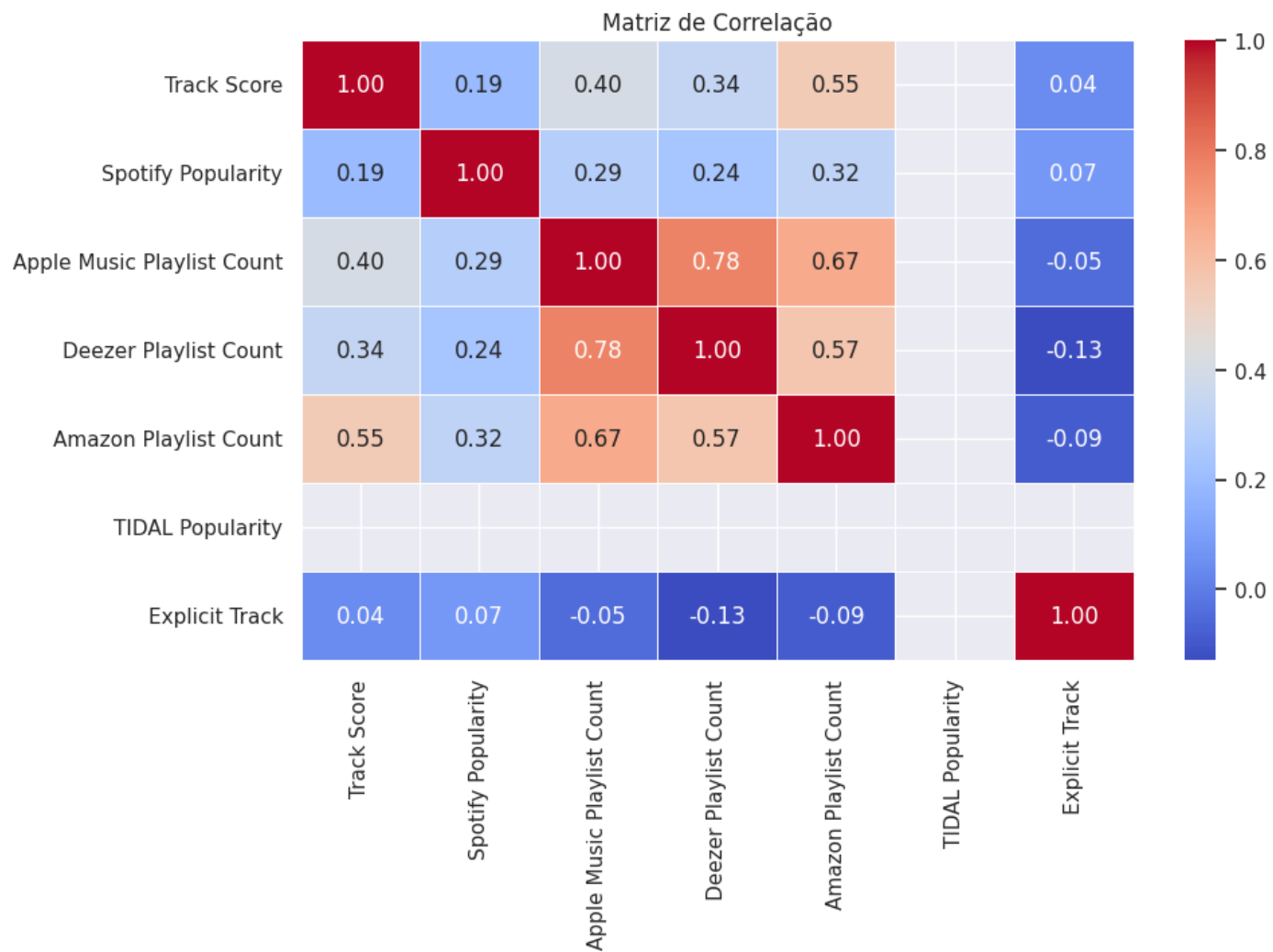


```
In [16]: # Distribuição das variáveis numéricas
df[numerical_cols].hist(figsize=(12, 10), bins=20, edgecolor='black')
plt.suptitle("Distribuição das variáveis numéricas")
plt.show()
```

Distribuição das variáveis numéricas

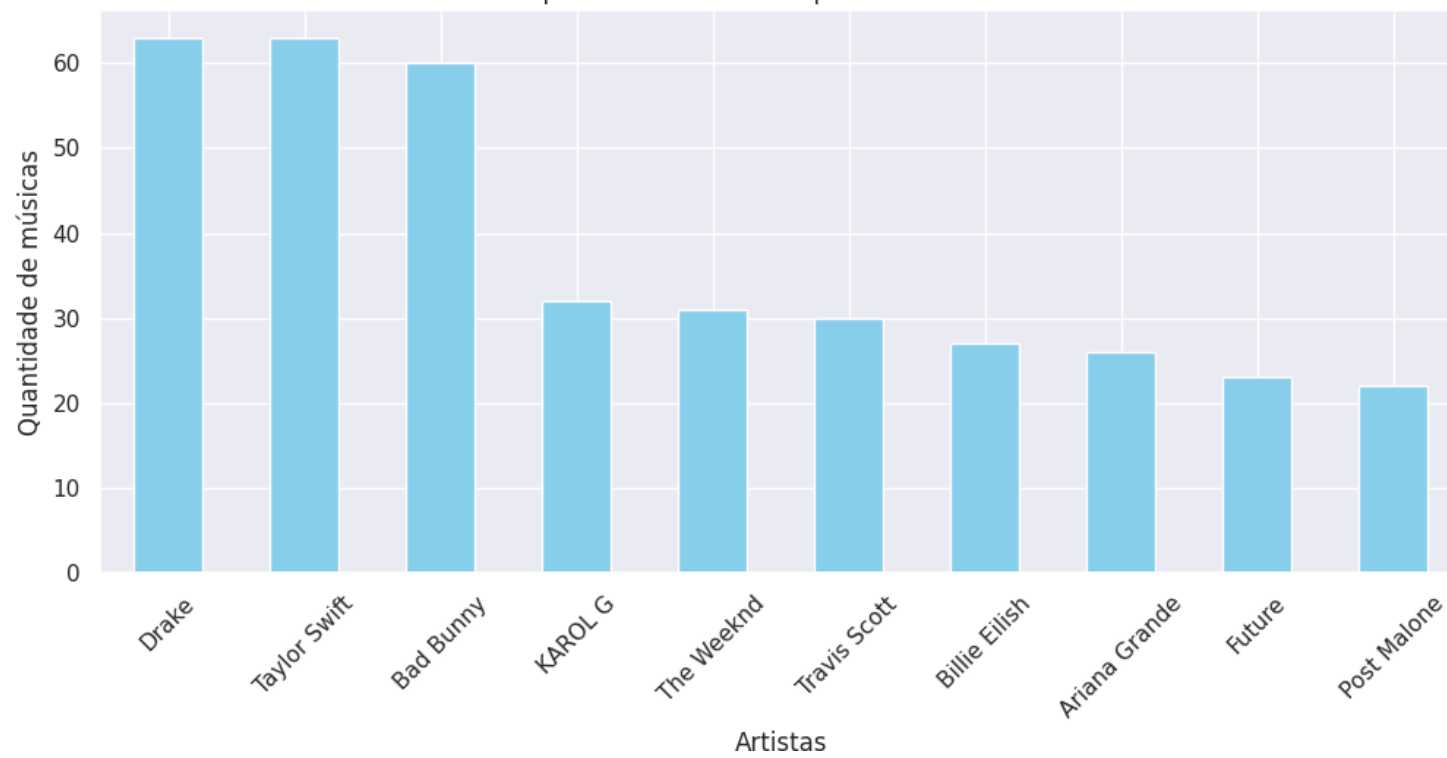


```
In [17]: # Matriz de correlação
plt.figure(figsize=(10, 6))
sns.heatmap(df[numerical_cols].corr(), annot=True, cmap="coolwarm", fmt=".2f", linewidths=0.5)
plt.title("Matriz de Correlação")
plt.show()
```

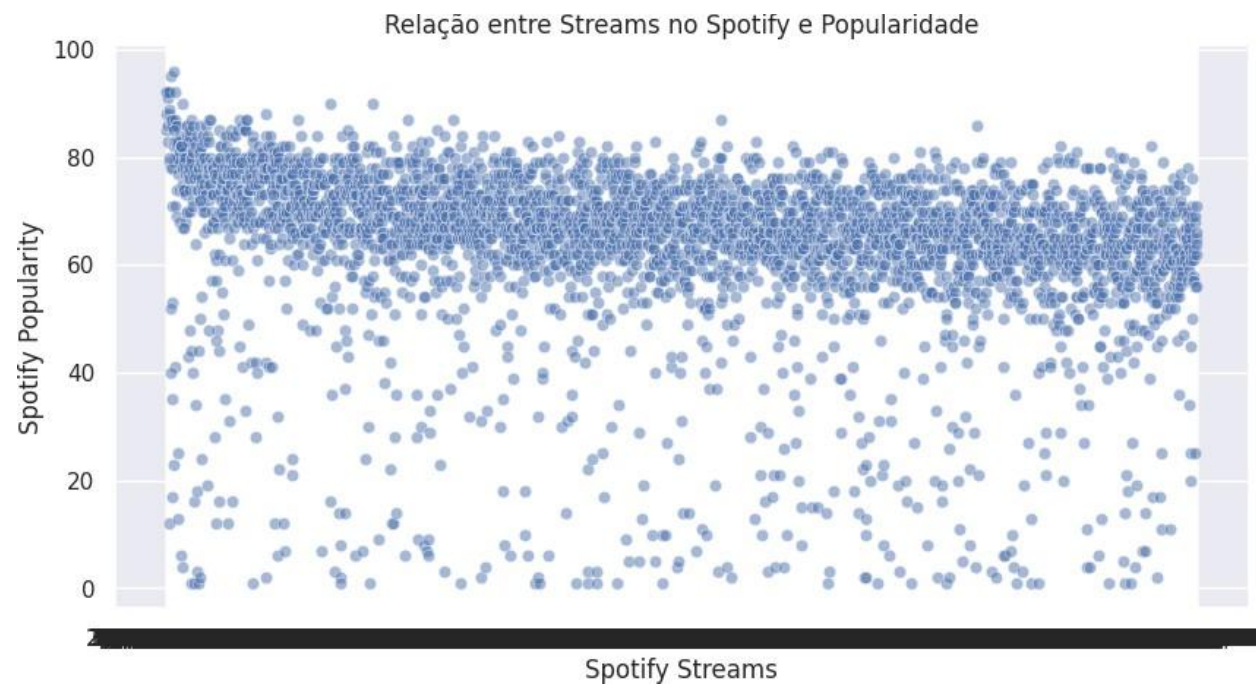


```
In [20]: # Top 10 artistas mais presentes
plt.figure(figsize=(12, 5))
df['Artist'].value_counts().head(10).plot(kind='bar', color='skyblue')
plt.title("Top 10 Artistas Mais Frequentes no Dataset")
plt.ylabel("Quantidade de músicas")
plt.xlabel("Artistas")
plt.xticks(rotation=45)
plt.show()
```

Top 10 Artistas Mais Frequentes no Dataset



```
In [28]: # Relação entre Popularidade e Streams do Spotify
plt.figure(figsize=(10, 5))
sns.scatterplot(data=df, x='Spotify Streams', y='Spotify Popularity', alpha=0.5)
plt.title("Relação entre Streams no Spotify e Popularidade")
plt.xlabel("Spotify Streams")
plt.ylabel("Spotify Popularity")
plt.show()
```



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

A análise dos dados das músicas mais reproduzidas no Spotify em 2024 permite compreender as tendências e preferências do público na era do streaming. A partir do dataset disponível no Kaggle, será possível identificar padrões que influenciam o sucesso de uma faixa, incluindo gêneros musicais, artistas e gravadoras mais populares. Esses insights são valiosos tanto para a indústria musical quanto para pesquisadores e profissionais de dados. Assim, espera-se que, o estudo contribua para uma visão mais ampla sobre o impacto do streaming no consumo de música e na definição do sucesso musical.