



**UNIVERSIDADE FEDERAL DO CEARÁ - CAMPUS CRATEÚS
CIÊNCIA DOS DADOS**

RELATÓRIO:

[T2] Análise Exploratória de Dados e Visualizações

DISCENTES:

ALYCIA ALVES ANDRADE - 536593

ANNA IWINY ALVES ANDRADE - 540134

JOSE WYTALO ADRIANO PACIFICO - 542305

WILDNEY KESNEY RODRIGUES DE SOUSA - 535767

CRATEÚS

2025

SUMÁRIO

1. INTRODUÇÃO.....	3
2. DADOS UTILIZADOS.....	3
3. OBJETIVO.....	4
4. PRINCIPAIS ACHADOS.....	4
4.1) Possível melhor dia para realizar lançamento de músicas:.....	4
4.2) Percepção acerca do ano de lançamento das vinte músicas mais ouvidas em 2023:.....	4
4.3) Os vinte artistas mais ouvidos em 2023 não necessariamente tem alguma música emplacada na lista das 20 músicas mais ouvidas.....	5
5. DESAFIOS E APRENDIZADOS.....	6
5.1) Desafios encontrados:.....	6
5.2) Aprendizados adquiridos:.....	6

1. INTRODUÇÃO

O presente trabalho está fundamentado a partir dos conhecimentos adquiridos na disciplina de Ciência dos Dados, no qual os integrantes buscaram aplicar conceitos de Análise Exploratória e Visualização de Dados. Para tanto, uma base de dados foi selecionada a fim de investigar suas informações relacionadas, bem como compreender sua estrutura, padrões e possíveis tendências presentes neste conjunto de dados, demonstrados por meios de métodos estatísticos (*plotação de gráficos*).

2. DADOS UTILIZADOS

A base de dados selecionada para a realização do trabalho é intitulada “**Most Streamed Spotify Songs 2023**”, disponibilizada na plataforma on-line e comunitária *Kaggle*, pelo usuário Nidula Elgiriyewithana. O conjunto de dados recebeu sua última atualização há cerca de dois anos, e nele estão contidas informações detalhadas sobre as músicas mais populares do ano de 2023 presentes nas plataformas de streaming, como Spotify, Deezer e Apple Music.

Além disso, no que diz respeito ao seu formato, trata-se de um arquivo do tipo CSV, ideal para análise de dados tabulares. Quanto ao seu volume, abrange um total de 953 músicas distintas.

E para além dos recursos que são normalmente encontrados em bases de dados semelhantes, esse dataset tem características que permitem realizar análises profundas sobre os atributos das músicas, sua popularidade e presença nas plataformas estudadas. Nesse contexto, destacam-se os seguintes atributos:

- **track_name**: nome da música;
- **artist(s)_name**: nome dos artistas responsáveis pela música;
- **released_date**: data de lançamento das músicas;
- **streams**: número de vezes que essa música foi tocada no Spotify;
- **audio features**: atributos que dizem respeito ao áudio das músicas, como duração, tonalidade, sua energia entre outros.

Os dados de catálogo musical (metadados de faixas, artistas e métricas de áudio) obtidos, ainda que publicamente, está estritamente sujeita aos Termos de Desenvolvedor do Spotify. Contudo, para projetos de análise com fins educacionais ou não comerciais, o uso dos dados é geralmente permitido, desde que a marca "Spotify" não seja utilizada indevidamente e não haja tentativa de reprodução do áudio.

Sobre éticas e de privacidade, como este *dataset* é composto por dados agregados e métricas públicas (como streams, danceability, ano de lançamento, etc.), as preocupações de privacidade pessoal são significativamente menores, pois o foco da análise recai sobre o conteúdo e as tendências do mercado, e não sobre a identificação ou o comportamento individual de usuários.

3. OBJETIVO

O objetivo deste trabalho é realizar uma Análise Exploratória de Dados, fundamentada no dataset escolhido. Dessa forma é possível identificar padrões de popularidade de músicas e artistas no ano de 2023, além de entender como se deram os lançamentos e seu desempenho. A partir dessa análise, é possível gerar *insights* por meio da visualização de dados, que auxiliam na compreensão dos possíveis impactos sociais, financeiros e estratégicos, causados por esses dados, sejam eles contextualizados na época estudada ou nos próximos anos.

4. PRINCIPAIS ACHADOS

4.1) Possível melhor dia para realizar lançamento de músicas:

Essa percepção surge a partir da análise dos dias de lançamentos das vinte músicas mais bem-sucedidas no ano de 2023:

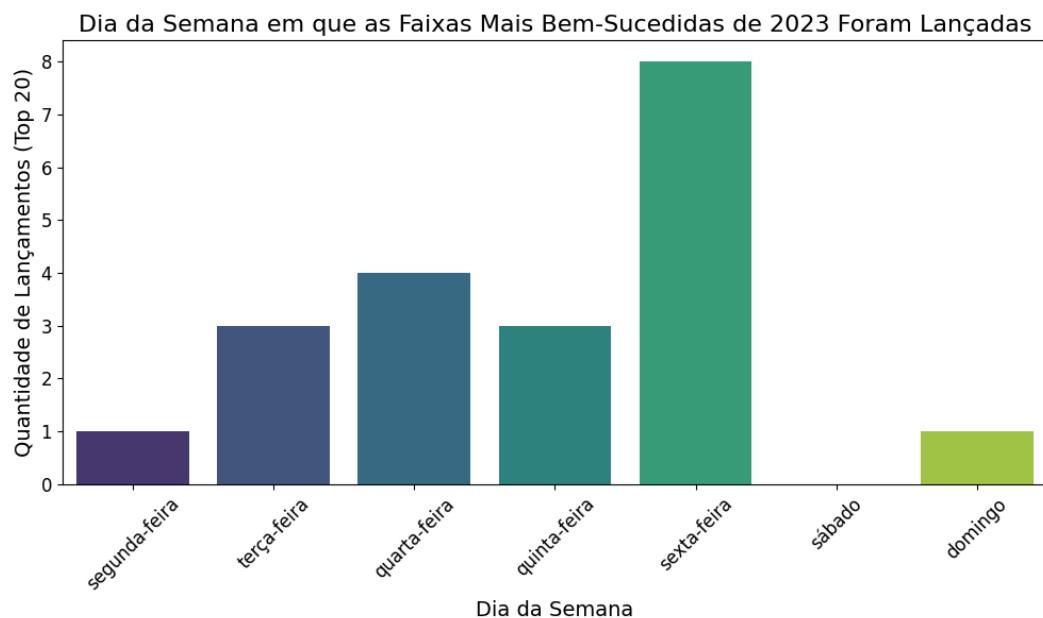


Figura 1: Gráfico de autoria própria

4.2) Percepção acerca do ano de lançamento das vinte músicas mais ouvidas em 2023:

Essa percepção fundamenta-se a partir da análise anterior. Para além dos dias da semana em que as músicas anteriores haviam sido lançadas, havia variação no ano de lançamento. Além disso, nenhuma dessas músicas foi lançada no ano de estudo, 2023, como demonstrado na figura abaixo:

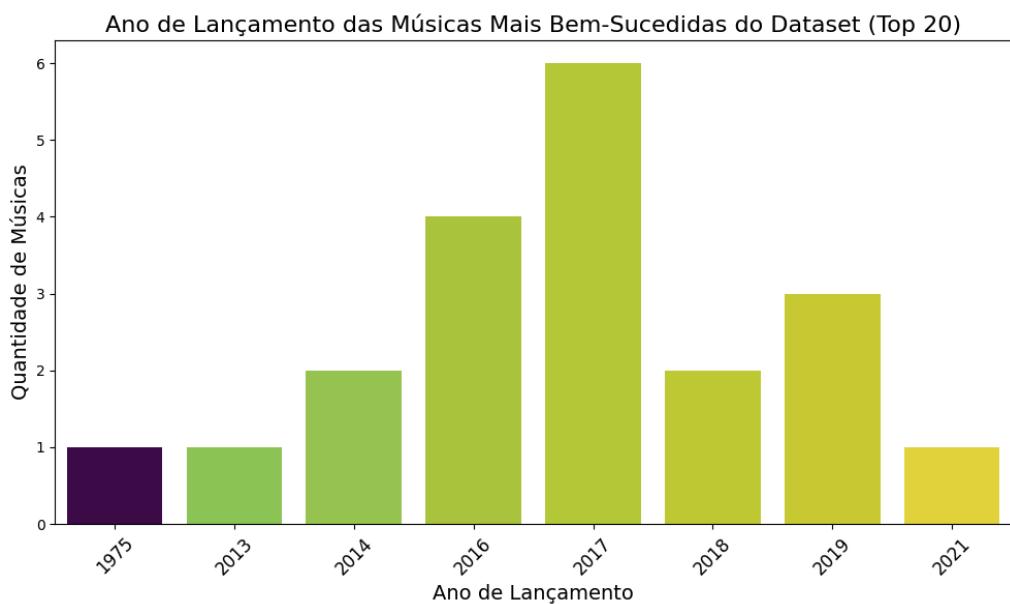


Figura 2: Gráfico de autoria própria

4.3) Os vinte artistas mais ouvidos em 2023 não necessariamente tem alguma música emplacada na lista das 20 músicas mais ouvidas.

A análise de dados para este caso permite identificar, para além da popularidade geral dos artistas mais ouvidos, o seu impacto direto na parada de músicas mais ouvidas do ano de 2023.

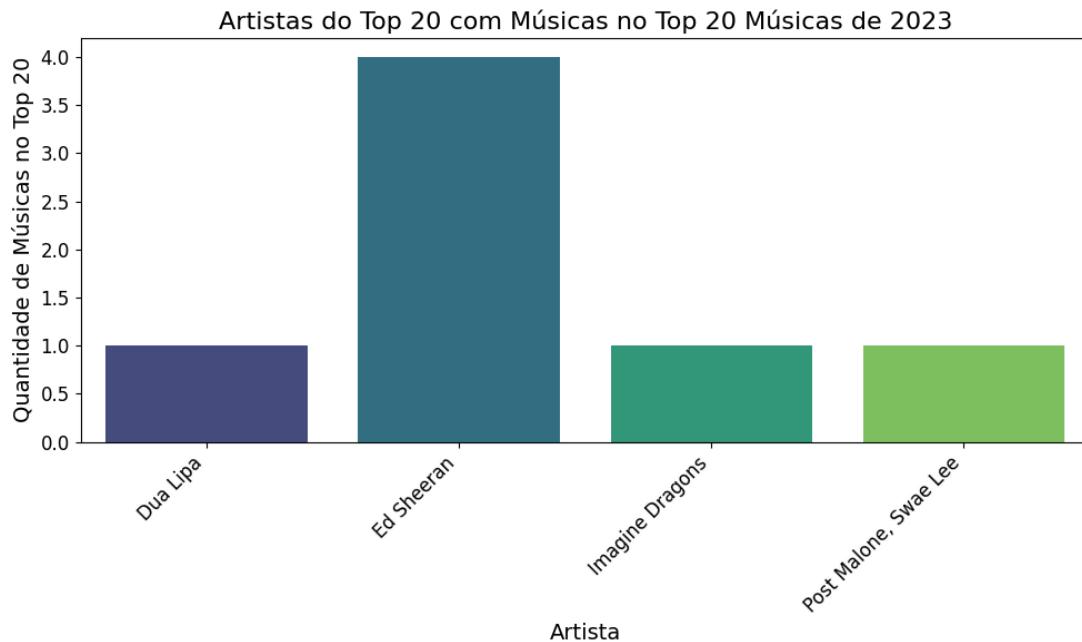


Figura 3: Gráfico de autoria própria

5. DESAFIOS E APRENDIZADOS

5.1) Desafios encontrados:

Grandezas Numéricas nos Gráficos: Devido à grande variação nos valores de streams, o matplotlib passou a exibir automaticamente os eixos em notação científica, o que prejudicava a leitura dos gráficos. Para resolver isso, criamos uma função de formatação que converte esses valores para unidades mais intuitivas (mil, milhões e bilhões), tornando as visualizações mais claras e fáceis de interpretar.

Limpeza de Dados e Datas: Foi necessário converter a coluna streams para o tipo numérico, remover caracteres de vírgula e padronizar formatos da Release Date para garantir a precisão dos cálculos e a execução dos gráficos.

Melhor Gráfico para Cada Pergunta: Evitamos o excesso de gráficos de barra, utilizando o Gráfico de Dispersão (*Regplot*) para correlação e o Gráfico de Caixa (*Box Plot*) para comparar distribuições de variáveis categóricas.

Perguntas e seus Propósitos: Pensar em questionamentos que retornem determinado benefício, sejam eles sociais ou financeiros, bem como as decisões que podem ser tomadas a partir disso, se tornou um desafio.

5.2) Aprendizados adquiridos:

Entendimento de Visualização: Aprendemos a adaptar a forma de mostrar os dados para garantir que a informação ficasse clara, o que nos permitiu visualizar e compreender melhor o *dataset* e a verdadeira performance das músicas.

Limpeza de Dados: O projeto destacou que a limpeza de dados é a base de qualquer análise confiável. Aprendemos que uma boa prática é sempre verificar e tratar os dados faltantes ou incorretos nas colunas importantes, pois essa etapa simples garante a precisão de todos os gráficos e cálculos feitos em seguida.