

# Atividade 02 - Análise exploratória

## Equipe

**Nome:** Donda

**Repositório GitLab da equipe:** [https://gitlab.com/jr\\_waine/csb-53](https://gitlab.com/jr_waine/csb-53)

**Membros:**

Waine Barbosa de Oliveira Junior, 1905120,  
jr\_waine, [waine@alunos.utfpr.edu.br](mailto:waine@alunos.utfpr.edu.br), Eng. Comp., UTFPR  
Eduardo Yoshio da Rocha, 1508733,  
eduardo2798, [eduardo.yoshio@outlook.com](mailto:eduardo.yoshio@outlook.com), Eng. Comp., UTFPR

## Tema

Analisar os fatores determinantes para indicação de melhor álbum no Grammy Awards.

## Motivação

É muito comum ouvir críticas acerca do Grammy e suas indicações para melhor música e álbum do ano, alegando que os álbuns são indicados não pela qualidade, mas sim porque foram populares ou sucesso de vendas.

Uma análise criteriosa dos dados de álbuns, como nota da crítica, número de vendas/ouvintes do álbum, deve ajudar a checar os argumentos dessa discussão e seu respaldo na realidade.

## Obtenção e processamento de dados

Para as métricas necessárias, coletamos dados de 3 fontes diferentes, [Billboard](#), [Metacritic](#) e dos álbuns indicados ao [Grammy](#). Para todas as fontes foram utilizados os anos de 2000 até 2020.

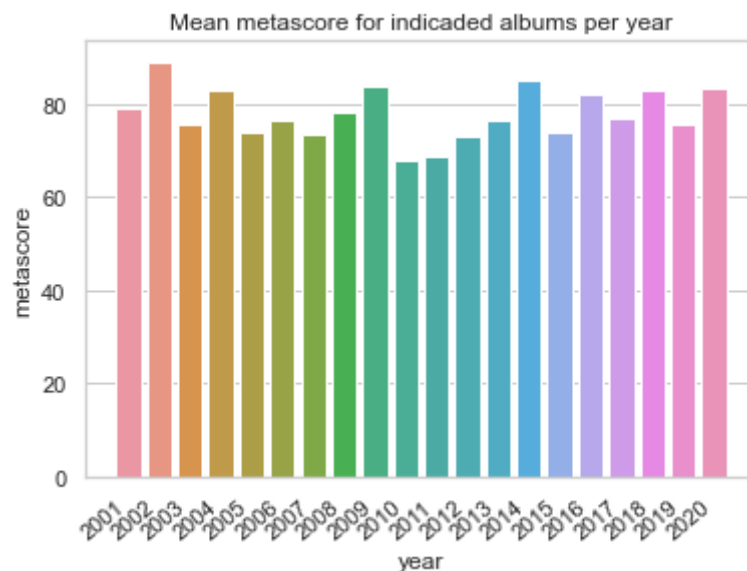
Na [Billboard](#) buscamos os top 200 álbuns mais vendidos dos EUA no ano. Para isso, foram baixadas as páginas e feito scrapping delas.

No [Metacritic](#) buscamos os top 100 álbuns melhores avaliados do ano e também os álbuns indicados ao Grammy. Para isso, foram baixadas as páginas necessárias e feito o parsing delas. Os dados obtidos foram: artista; álbum; data de lançamento; *userscore* (nota média dos usuários); *metascore* (nota média da crítica); link do álbum no site; link da imagem do álbum no site.

Para o [Grammy](#), os dados foram obtidos manualmente, sendo obtido o ano, artista, álbum e se o álbum venceu ou não. Foi feito um parsing e um tratamento dos álbuns para obter suas informações do [Metacritic](#), sendo que alguns álbuns não foram encontrados, portanto não devem ser utilizados na análise das notas da crítica.

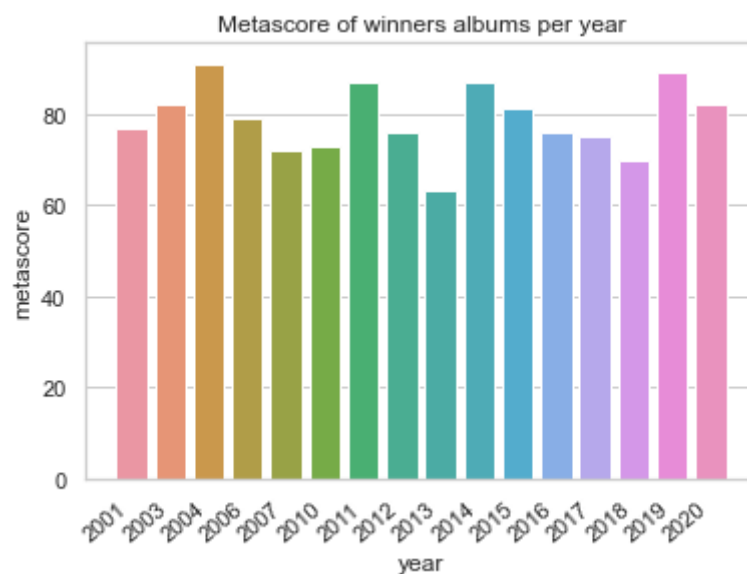
# Cobertura e distribuição dos dados

## *Metascore* médio dos indicados ao Grammy



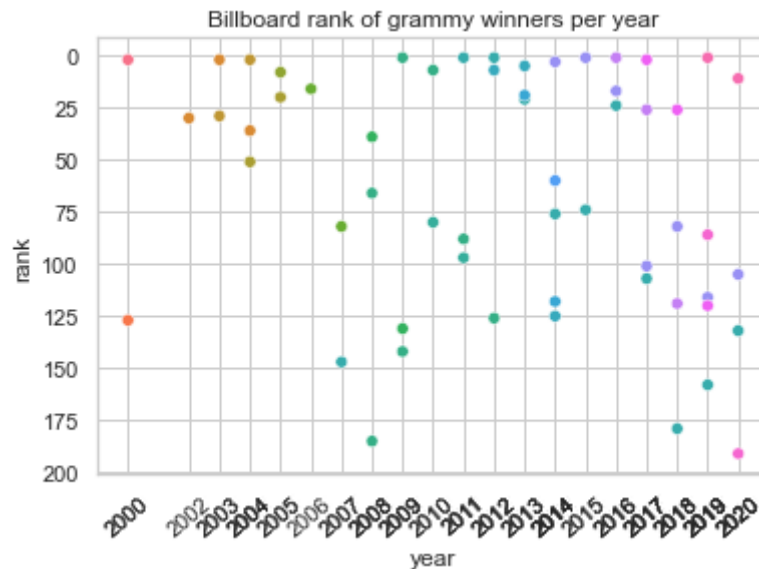
O metascore médio por ano dos indicados ao Grammy tem uma variação relativamente alta, com uma média de todos os anos de pouco abaixo de 80, considerados álbuns de qualidade para o Metacritic, mas não necessariamente de excelência. Vale notar os anos de 2011 até 2013 em que as médias tiveram os valores mais baixos de toda série.

## *Metascore* dos vencedores do Grammy



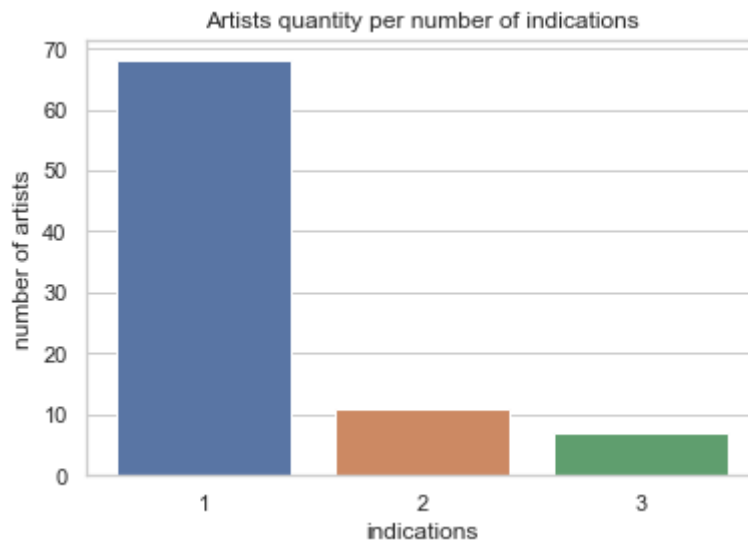
O gráfico mostra que os álbuns vencedores tem uma nota média pouco abaixo de 80, o que pode ser considerado um álbum de qualidade com relação à média do Metacritic, porém não de excelência.

## Rank na Billboard dos vencedores do Grammy por ano



É possível perceber que há uma concentração de álbuns nas primeiras posições em todo ano. Isso demonstra que há uma grande relação entre vencer o Grammy e ser popular. Há repetições no gráfico, pois os álbuns vencedores podem ficar mais de um ano no top 200, ou ser lançado no ano anterior com relação ao Grammy que concorreram.

## Número de artistas com N indicações



O gráfico mostra que a grande maioria dos artistas tem apenas uma indicação ao Grammy, com apenas 11 artistas com 2 indicações, e 6 com três.

## Perguntas de pesquisa e explorações iniciais

Relembrando os objetivos definidos

1. Definir métricas para popularidade e qualidade/relevância de um álbum
2. Definição dos dados a serem coletados (relacionados às métricas)

3. Coleta de dados dos álbuns de cada ano (álbuns indicados e de maior relevância)
4. Construção do modelo para análise dos dados
5. Definir modo de apresentação e visualização dos dados
6. Concluir sobre os resultados obtidos

Com relação às métricas e os dados a serem coletados, utilizamos o [Metacritic](#) como referência para a qualidade técnica do álbum. Para popularidade utilizamos o top 200 [Billboard](#) em um curto intervalo, 1 a 2 anos, enquanto que para relevância, utilizamos o mesmo ranking num intervalo mais longo (+2 anos). Isso é feito pois há vários álbuns que foram lançados há muito tempo, alguns datando da década de 70, porém ainda hoje continuam relevantes no cenário musical.

Com relação à coleta dos dados, consideramos que foi feita com sucesso e temos a maior parte das informações necessárias para as próximas fases.

## Discussão e próximos passos

Consideramos que os dados básicos foram coletados e explorados com sucesso, porém ainda há dúvidas se são o bastante para a criação do modelo. Talvez seja necessário obter mais dados com relação aos álbuns para criação do modelo, como o gênero musical. Tendo isso em mente, já iniciamos a construção de um script para a obtenção desses dados.

Acreditamos também que é possível melhorar a visualização dos gráficos e seus estilos, porém as informações mais relevantes sobre os álbuns e seus dados foram apresentadas.