

Fatores para indicação de melhor álbum no Grammy Awards

Eduardo Yoshio da Rocha
Waine Oliveira Junior





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.



Objetivos

- Determinar características em comum de álbuns indicados ao Grammy
 - Popularidade, relevância, crítica positiva, gênero musical
- Semelhanças e diferenças entre álbuns indicados ao Grammy, aqueles aclamados pela crítica e os mais populares



Hipóteses

- Os álbuns indicados e vencedores do Grammy tem muito mais relação com sua popularidade que sua qualidade técnica (crítica positiva)
- Os gêneros dos álbuns indicados ao Grammy têm pouca diversidade com relação aos álbuns aclamados pela crítica



Dados

- Período de 2000 a 2020
- Indicados ao [Grammy](#)
- Top 200 [Billboard](#) para popularidade/relevância
- Top 100 [Metacritic](#) para crítica (metascore) e demais características dos álbuns (gênero, data lançamento, etc.)



Processo de obtenção dos dados

1. Álbuns indicados ao Grammy foram obtidos manualmente
2. Top 200 Billboard e top 100 metacritic foram obtidos a partir de web-scraping
3. Com todos os álbuns, um parser para URL do Metacritic do álbum foi gerado
4. As páginas dos álbuns (aquelas existentes) foram baixadas
5. Web-scraping para obtenção dos dados da página do álbum
6. Dados agrupados em um JSON

Obs.: álbuns que não foram encontrados no [Metacritic](#) não foram considerados nas análises

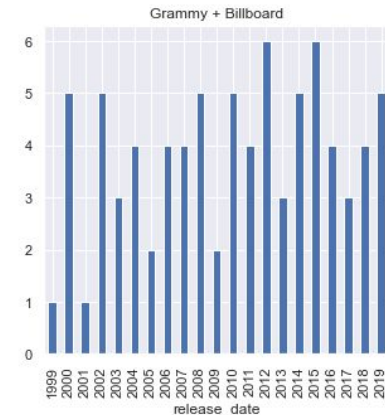
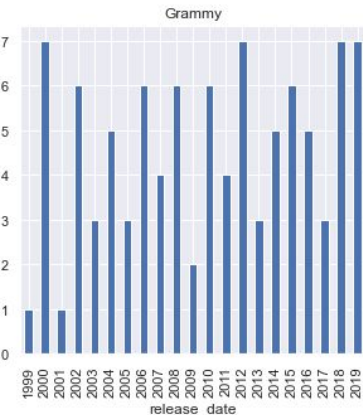
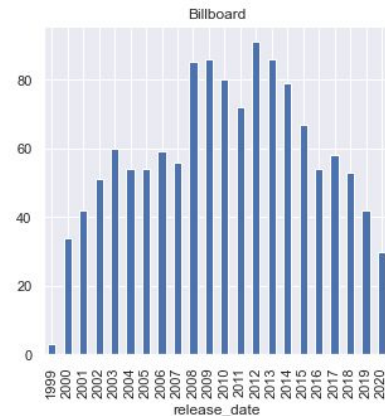
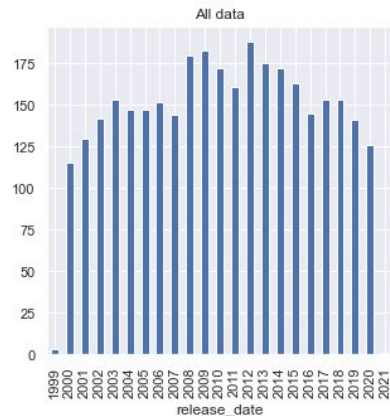


Dados

Billboard: 200 por ano é o ideal

Grammy: ~5 por ano é o ideal

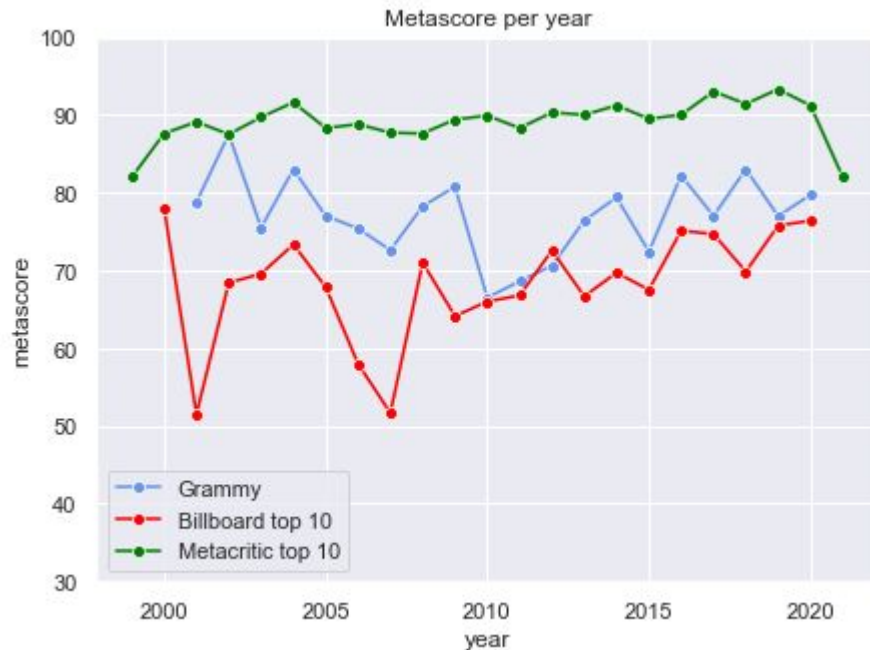
O número de álbuns utilizados é satisfatório
para nossas análises



Grammy x Billboard x Metacritic (temporal)

Os indicados ao Grammy tem um metaspcore melhor que o top 10 Billboard, na série histórica.

Uma leve tendência de aumento
tendência no metaspcore dos álbuns do
Grammy e da Billboard a partir de 2010



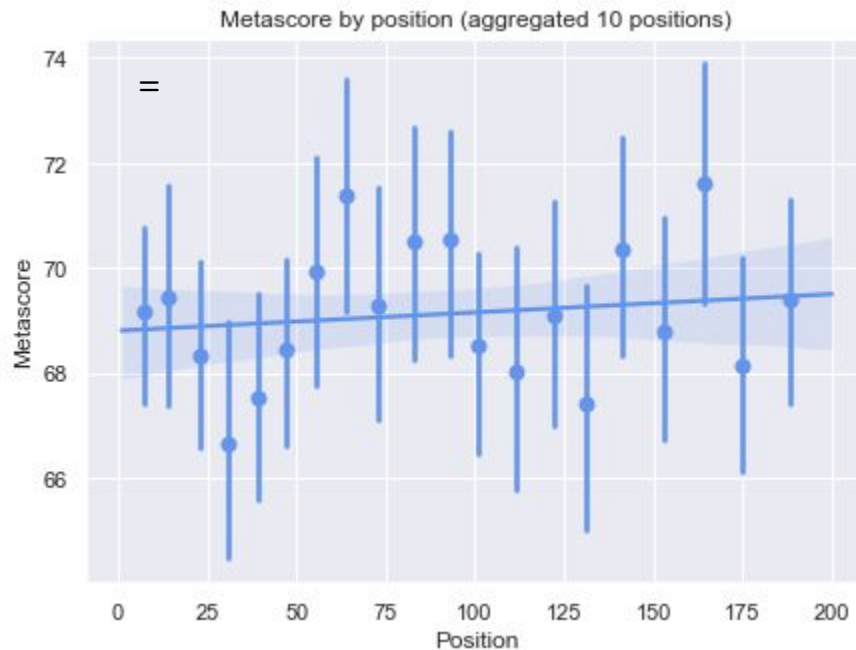


Billboard (posição)

Não é possível observar uma relação entre a posição na Billboard com o metascore.

$R=0.019$

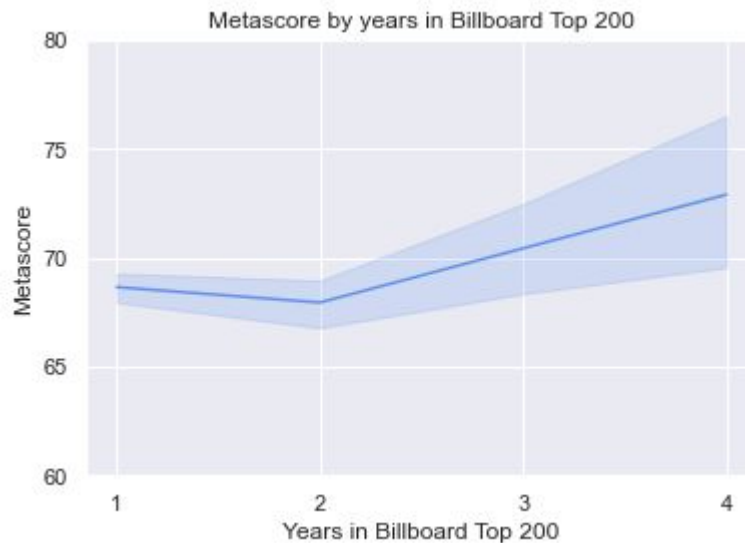
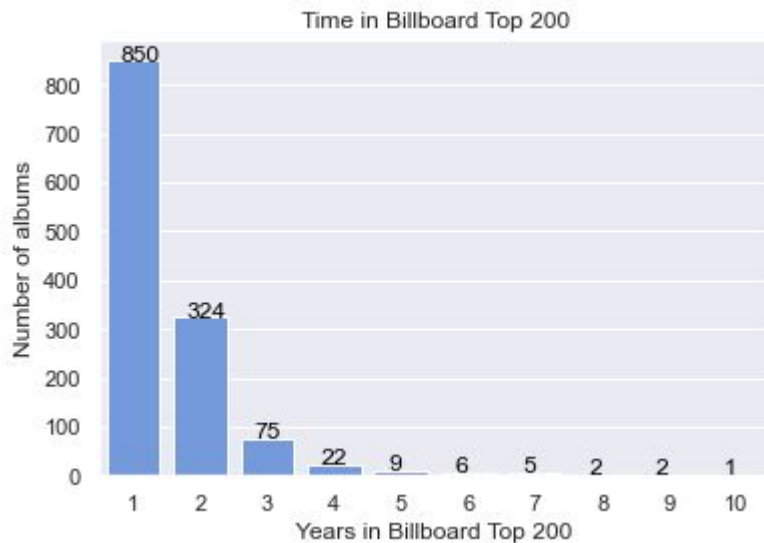
$P=0.40$





Billboard (nº de anos)

Há um leve aumento na média do metascore conforme o número de anos que o álbum fica no top 200 Billboard.

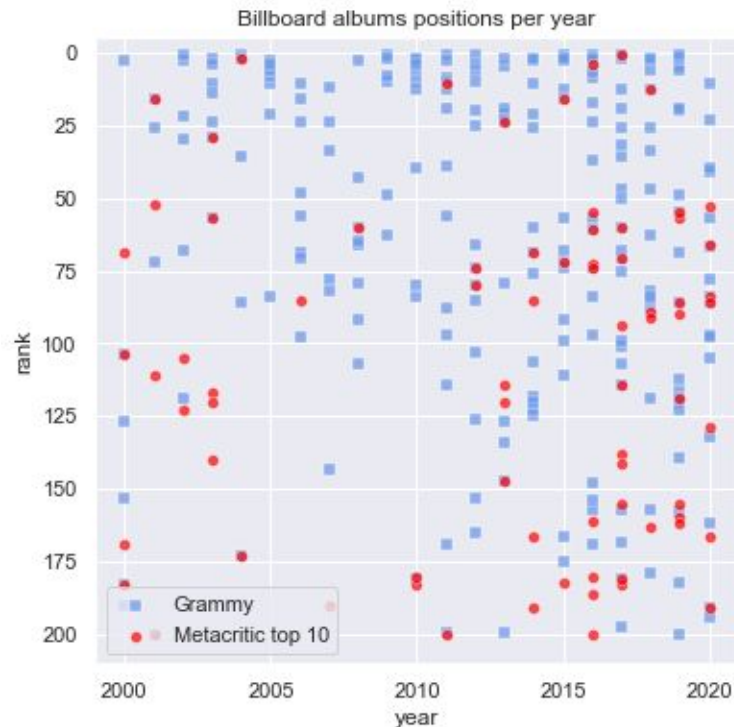




Posição na Billboard

Há muito mais álbuns indicados ao Grammy no Billboard top 200 do que do top 10 Metacritic.

Suspeitamos que a popularidade tem um grande impacto na indicação ao Grammy, ou vice-versa.

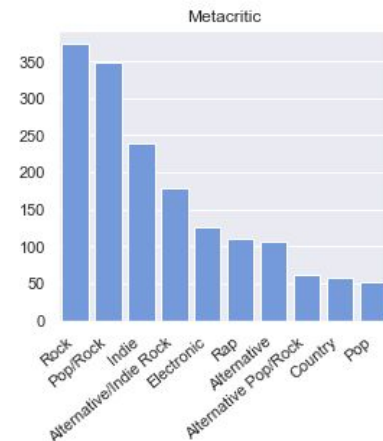
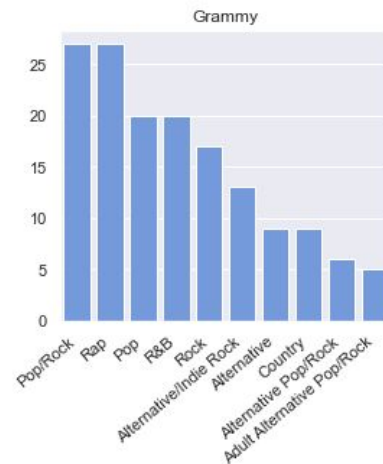
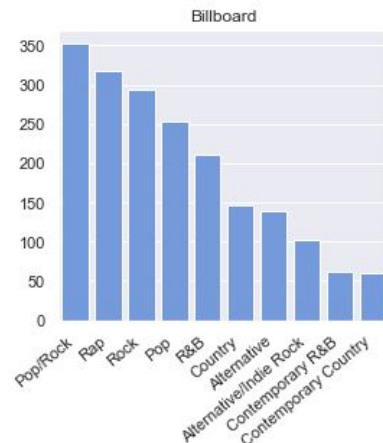
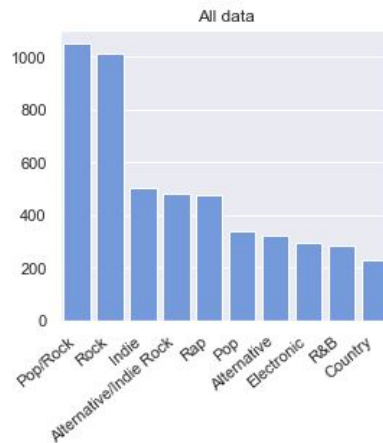




Gêneros

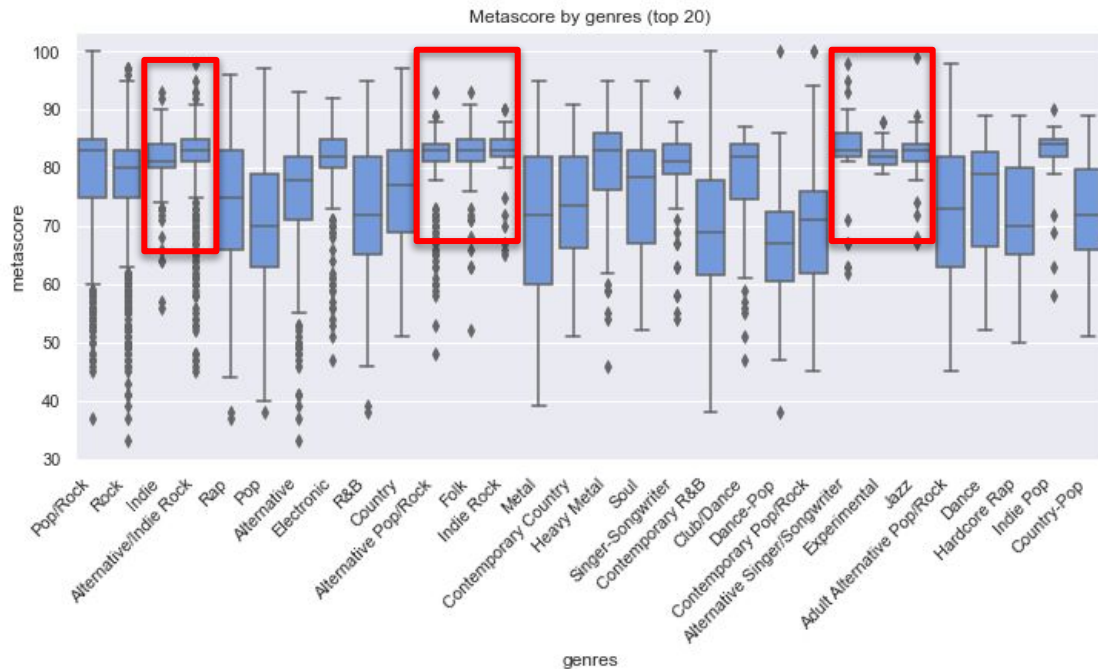
Gêneros mais comuns por dataset

- Pop/Rock, Pop, R&B e Rap se destacam em todos
- Os gênero predominantes de todos datasets são semelhantes, porém a distribuição/perfil diferem

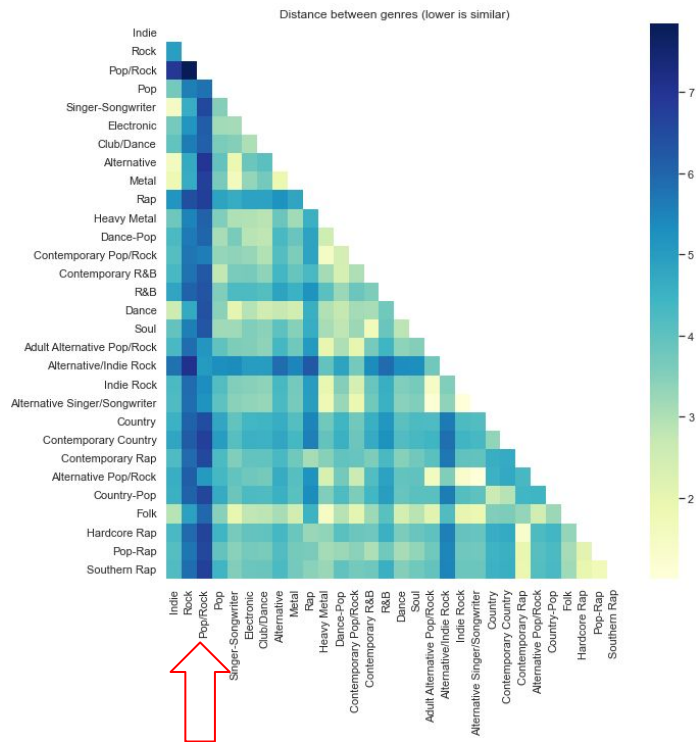


Gêneros

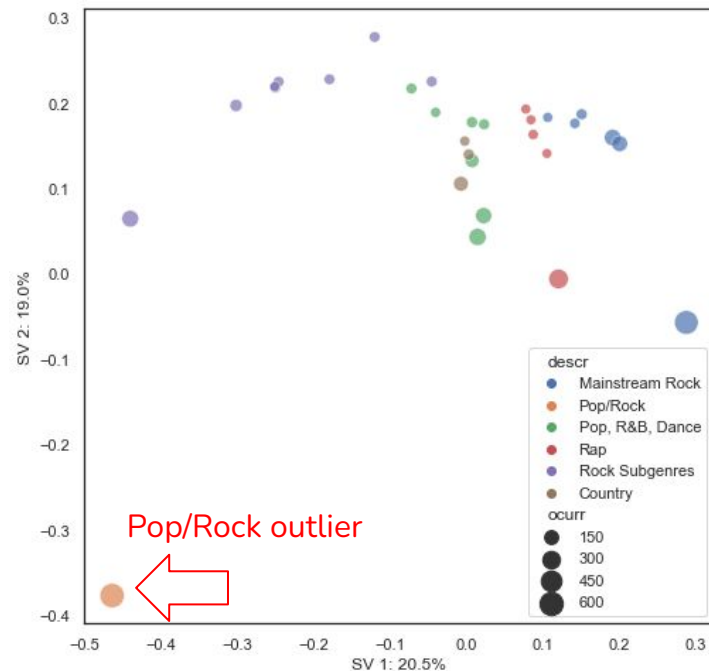
Gêneros com maior número de álbuns no Metacritic apresentam uma distribuição concentrada em notas altas



Gêneros



Utilizados apenas álbuns do Grammy, Billboard ou top 50 Metacritic para balancear fontes



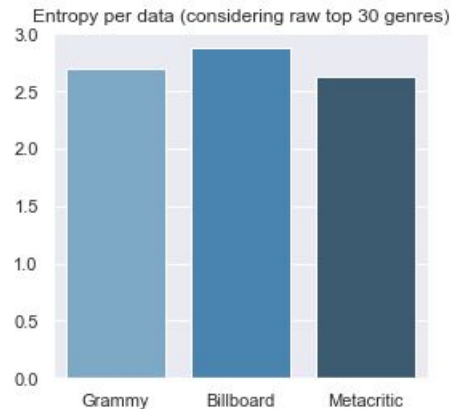
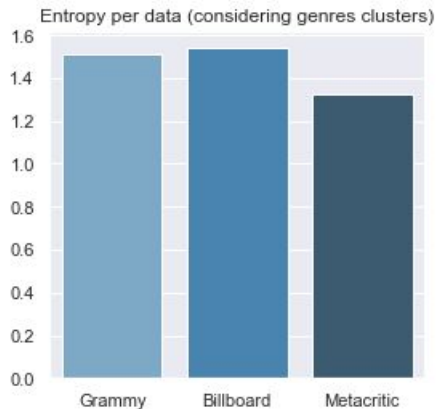


Entropia de Gêneros

Maior entropia quer dizer mais aleatório os gêneros.

$$H(X) = - \sum_{i=1}^n P(x_i) \log P(x_i)$$

- Grammy e Billboard com entropia semelhante nos dois casos
- Metacritic com a menor entropia (menos diversidade) nos dois casos





Fatores para indicação ao Grammy

- Foram considerados primeiramente 3 fatores para a indicação dos álbuns ao Grammy: Nota do Metacritic (Metascore), posição na Billboard e mês de lançamento.
- Analisando cada fator individualmente, temos:

Metascore

Logit Regression Results

Dep. Variable: grammy_indication

No. Observations: 3246

Model: Logit

Df Residuals: 3244

Method: MLE

Df Model: 1

Date: Sat, 11 Dec 2021

Pseudo R-squ.: 4.373e-05

Time: 19:08:32

Log-Likelihood: -436.03

converged: True

LL-Null: -436.05

Covariance Type: nonrobust

LLR p-value: 0.8452

coef

std err

z

P>|z|

[0.025

0.975]

Intercept

-3.3259

0.793

-4.196

0.000

-4.879

-1.773

metascore

-0.0020

0.010

-0.196

0.845

-0.022

0.018



Fatores para indicação ao Grammy

- Foram considerados primeiramente 3 fatores para a indicação dos álbuns ao Grammy: Nota do Metacritic (Metascore), posição na Billboard e mês de lançamento.
- Analisando cada fator individualmente, temos ao fazer a regressão logística o seguinte:

Metascore

| Logit Regression Results | | | | | | |
|--------------------------|-------------------|-------------------|-----------|-------|--------|--------|
| Dep. Variable: | grammy_indication | No. Observations: | 3246 | | | |
| Model: | Logit | Df Residuals: | 3244 | | | |
| Method: | MLE | Df Model: | 1 | | | |
| Date: | Sat, 11 Dec 2021 | Pseudo R-squ.: | 4.373e-05 | | | |
| Time: | 19:08:32 | Log-Likelihood: | -436.03 | | | |
| converged: | True | LL-Null: | -436.05 | | | |
| Covariance Type: | nonrobust | LLR p-value: | 0.8452 | | | |
| | coef | std err | z | P> z | [0.025 | 0.975] |
| Intercept | -3.3259 | 0.793 | -4.196 | 0.000 | -4.879 | -1.773 |
| metascore | -0.0020 | 0.010 | -0.196 | 0.845 | -0.022 | 0.018 |

Fatores para indicação ao Grammy

Logit Regression Results

| | | | |
|------------------|-------------------|-------------------|-----------|
| Dep. Variable: | grammy_indication | No. Observations: | 3246 |
| Model: | Logit | Df Residuals: | 3244 |
| Method: | MLE | Df Model: | 1 |
| Date: | Sat, 11 Dec 2021 | Pseudo R-squ.: | 0.1785 |
| Time: | 19:08:32 | Log-Likelihood: | -358.23 |
| converged: | True | LL-Null: | -436.05 |
| Covariance Type: | nonrobust | LLR p-value: | 1.016e-35 |

| | coef | std err | z | P> z | [0.025 | 0.975] |
|---------------------|---------|---------|---------|-------|--------|--------|
| Intercept | -1.5577 | 0.150 | -10.395 | 0.000 | -1.851 | -1.264 |
| best_rank_billboard | -0.0176 | 0.002 | -10.909 | 0.000 | -0.021 | -0.014 |

Posição na
Billboard (melhor)



Mês de Lançamento



Logit Regression Results

| | | | |
|------------------|-------------------|-------------------|----------|
| Dep. Variable: | grammy_indication | No. Observations: | 3246 |
| Model: | Logit | Df Residuals: | 3244 |
| Method: | MLE | Df Model: | 1 |
| Date: | Sat, 11 Dec 2021 | Pseudo R-squ.: | 0.001169 |
| Time: | 19:08:32 | Log-Likelihood: | -435.54 |
| converged: | True | LL-Null: | -436.05 |
| Covariance Type: | nonrobust | LLR p-value: | 0.3126 |

| | coef | std err | z | P> z | [0.025 | 0.975] |
|---------------|---------|---------|---------|-------|--------|--------|
| Intercept | -3.7011 | 0.247 | -14.985 | 0.000 | -4.185 | -3.217 |
| release_month | 0.0327 | 0.032 | 1.007 | 0.314 | -0.031 | 0.096 |



Fatores para indicação ao Grammy

Considerando todos os fatores:

- Metascore torna-se significativo
- Mês de lançamento ainda sem significância
- R ainda baixo
- Possível interação entre Metascore e Posição na Billboard

Logit Regression Results

```
=====
Dep. Variable:    grammy_indication    No. Observations:    3246
Model:            Logit                Df Residuals:        3242
Method:           MLE                  Df Model:            3
Date:            Sat, 11 Dec 2021      Pseudo R-squ.:      0.2369
Time:            19:08:32              Log-Likelihood:     -332.73
converged:        True                  LL-Null:            -436.05
Covariance Type:  nonrobust             LLR p-value:        1.554e-44
=====
```

```
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
Intercept      -7.4348      0.986     -7.538      0.000     -9.368     -5.502
metascore        0.0844      0.013      6.622      0.000      0.059      0.109
best_rank_billboard -0.0223      0.002    -13.118      0.000     -0.026     -0.019
release_month   -0.0049      0.033     -0.148      0.882     -0.070      0.060
=====
```

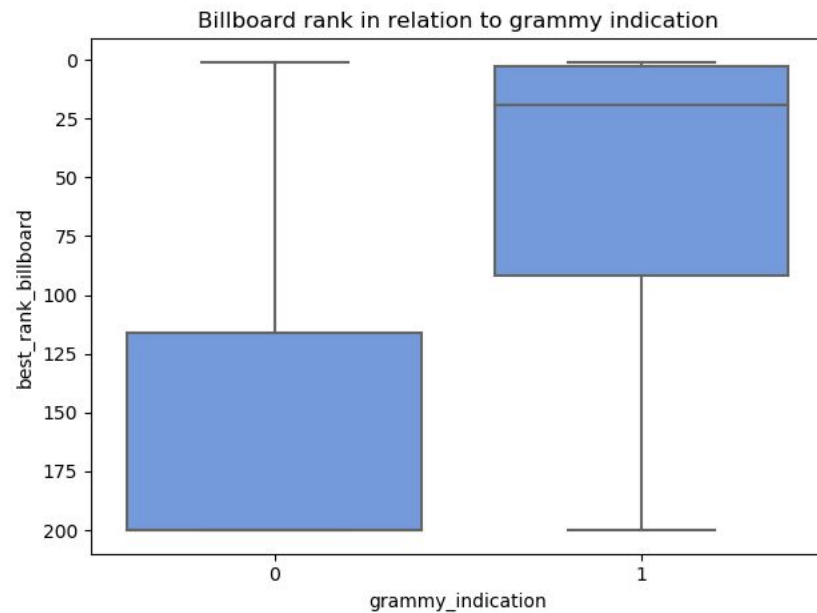
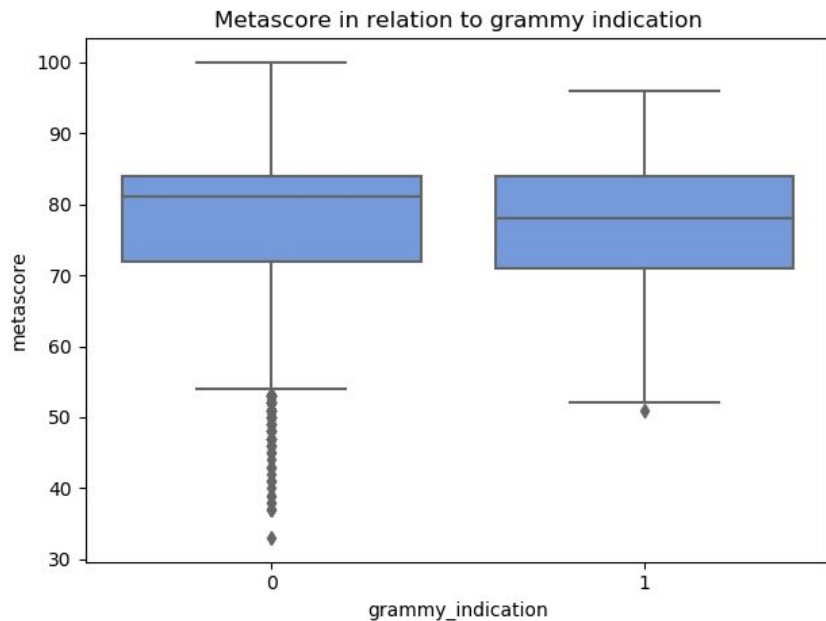
Odds Ratio



| | 5% | 95% | Odds Ratio |
|---------------------|----------|----------|------------|
| Intercept | 0.000085 | 0.004080 | 0.000590 |
| metascore | 1.061257 | 1.115655 | 1.088116 |
| best_rank_billboard | 0.974676 | 0.981197 | 0.977931 |
| release_month | 0.932664 | 1.061739 | 0.995111 |

Fatores para indicação ao Grammy

- Posição na Billboard tem maior influência na indicação ao grammy



Fatores para indicação ao Grammy

- Predição de indicação ao Grammy

```
metascore best_rank_billboard release_month
0         25                131             4
Chance to be indicated to grammy awards:
0    0.000257
```

```
metascore best_rank_billboard release_month
0         51                183             3
Chance to be indicated to grammy awards:
0    0.000727
```

```
metascore best_rank_billboard release_month
0         41                116             2
Chance to be indicated to grammy awards:
0    0.001399
```

```
metascore best_rank_billboard release_month
0         21                 19             11
Chance to be indicated to grammy awards:
0    0.002152
```

```
metascore best_rank_billboard release_month
0         43                 64             11
Chance to be indicated to grammy awards:
0    0.005038
```

```
metascore best_rank_billboard release_month
0         53                 33             5
Chance to be indicated to grammy awards:
0    0.023661
```

```
metascore best_rank_billboard release_month
0         89                 140            8
Chance to be indicated to grammy awards:
0    0.043841
```

```
metascore best_rank_billboard release_month
0         74                 24             10
Chance to be indicated to grammy awards:
0    0.145515
```

Fatores para indicação ao Grammy

- Regressão Logística considerando os gêneros musicais
- Valores de distância menores indicam álbuns pertencentes àqueles gêneros

| Logit Regression Results | | | | | | |
|--------------------------|-------------------|-------------------|-----------|-------|---------|--------|
| Dep. Variable: | grammy_indication | No. Observations: | 3155 | | | |
| Model: | Logit | Df Residuals: | 3145 | | | |
| Method: | MLE | Df Model: | 9 | | | |
| Date: | Sat, 11 Dec 2021 | Pseudo R-squ.: | 0.2400 | | | |
| Time: | 21:18:15 | Log-Likelihood: | -326.64 | | | |
| converged: | True | LL-Null: | -429.79 | | | |
| Covariance Type: | nonrobust | LLR p-value: | 1.579e-39 | | | |
| | coef | std err | z | P> z | [0.025 | 0.975] |
| Intercept | -9.7582 | 1.589 | -6.141 | 0.000 | -12.873 | -6.644 |
| metascore | 0.0874 | 0.013 | 6.589 | 0.000 | 0.061 | 0.113 |
| best_rank_billboard | -0.0229 | 0.002 | -12.294 | 0.000 | -0.027 | -0.019 |
| release_month | -0.0125 | 0.034 | -0.372 | 0.710 | -0.079 | 0.053 |
| Rap | 0.2064 | 0.138 | 1.499 | 0.134 | -0.063 | 0.476 |
| Pop_RnB_Dance | -0.0165 | 0.165 | -0.100 | 0.920 | -0.341 | 0.307 |
| Mainstream_Rock | 0.1805 | 0.142 | 1.268 | 0.205 | -0.098 | 0.459 |
| Country | 0.2028 | 0.129 | 1.573 | 0.116 | -0.050 | 0.455 |
| Rock_Subgenres | -0.2908 | 0.193 | -1.505 | 0.132 | -0.669 | 0.088 |
| Pop_Rock | 0.1866 | 0.111 | 1.688 | 0.091 | -0.030 | 0.403 |



Respostas às hipóteses

- Os álbuns indicados e vencedores do Grammy tem muito mais relação com sua popularidade que sua qualidade técnica (crítica positiva)?

Sim, a popularidade tem uma grande influência na indicação ao Grammy (ou vice-versa), maior que a avaliação crítica

- Os gêneros dos álbuns indicados ao Grammy têm pouca diversidade com relação aos álbuns aclamados pela crítica?

Não, na verdade a diversidade entre álbuns aclamados é menor que álbuns indicados ao Grammy ou populares (na Billboard)



Dificuldades

- Obtenção e processamento dos dados
- Pré-processamento com Pandas em estruturas como listas/dicionários em colunas
- Encontrar maneiras para testar as hipóteses
- Encontrar análises que trouxessem caminhos para investigarmos os fatores relevantes



Limitações

- Uma maior base de dados, com mais informações e maior intervalo de tempo, poderia permitir uma análise mais precisa
- A análise e clusterização de gênero foi feita apenas utilizando as ocorrências nos álbuns. Adicionar outras informações para essa clusterização permitiria uma melhor caracterização dos clusters.
- A modelagem de popularidade e de avaliação crítica dos álbuns é relativamente simples. Considerar outros fatores além da posição da Billboard e a nota no Metacritic poderia refletir melhor esses fatores.
- Há vários outros caminhos de análises a partir dos dados obtidos que podem ser explorados ainda (e.g. periodicidade de lançamento de álbuns por artista e influência disso nos fatores apresentados).