



# Proposta de novas playlists MozartBee

---





# Objetivo e Características Técnicas

**Objetivo:** A empresa MozartBee possui um novo e inovador aplicativo de música criado para ser concorrente do Spotify. Como o aplicativo busca novos ouvintes, a empresa precisa criar novas listas de músicas agrupando-as com base em suas características para sugerir novas playlists aos usuários.

**Tipo de modelo:** Não Supervisionado

**Técnica:** Clusterização

**Técnicas para definir clusters:** Silhouette e Elbow.

**Algoritmo:** KMeans

# Alinhamento com o negócio



- Quantidade de listas (cluster): no mínimo 4, a intenção é descobrir novas listas de indicação de músicas;
- Os artistas mais populares são os que possuem maiores streams;
- Variáveis consideradas: O dataset fornecido para análise já considera as colunas importantes para a regra de negócio



## Observações do cientista:

- Para a determinação da quantidade de Cluster iremos utilizar o método do cotovelo e o coeficiente de “Silhouette”, caso o valor seja menor do que o solicitado, iremos **adotar o mínimo** solicitado pelo time de negócios
- Iremos usar a distância euclidiana (k-means) para a fase de clusterização

# Metodologia – Análise exploratória



## Colunas e especificações:

Atributo	Descrição
track_name	Nome da música
artist(s)_name	Nome do(s) artista(s) que performa(m) a música
artist_count	Número de artistas contribuindo na música
released_year	Ano de lançamento
released_month	Mês de lançamento
released_day	Dia de lançamento
in_spotify_playlists	Número de playlists no Spotify onde a música está
in_spotify_charts	Posição da música nas paradas do Spotify
streams	Total de streams no Spotify
in_apple_playlists	Presença em playlists do Apple Music
in_apple_charts	Presença nas paradas do Apple Music
in_deezer_playlists	Presença em playlists do Deezer
in_deezer_charts	Presença nas paradas do Deezer
in_shazam_charts	Posição nas paradas do Shazam
bpm	Batidas por minuto (tempo da música)
key	Tom da música
mode	Modo maior ou menor
danceability_%	Adequação para dança (%)
valence_%	Positividade do conteúdo musical (%)
energy_%	Nível de energia percebido (%)
acousticness_%	Presença de som acústico (%)
instrumentalness_%	Proporção de conteúdo instrumental (%)
liveness_%	Presença de elementos de performance ao vivo (%)
speechiness_%	Quantidade de palavras faladas (%)

### Observações do cientista:

- Variáveis selecionadas: bpm, danceability\_%, energy\_%, valence\_%, liveness\_%, speechiness\_% para clusterização



## Como colunas foram definidas

- Describe

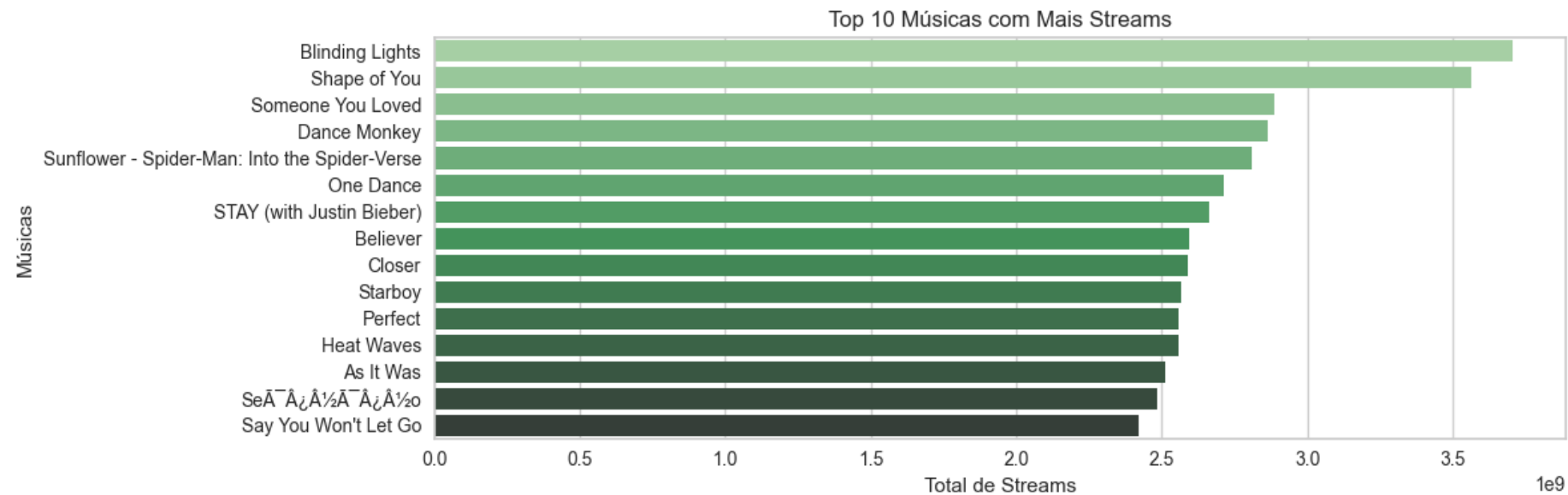
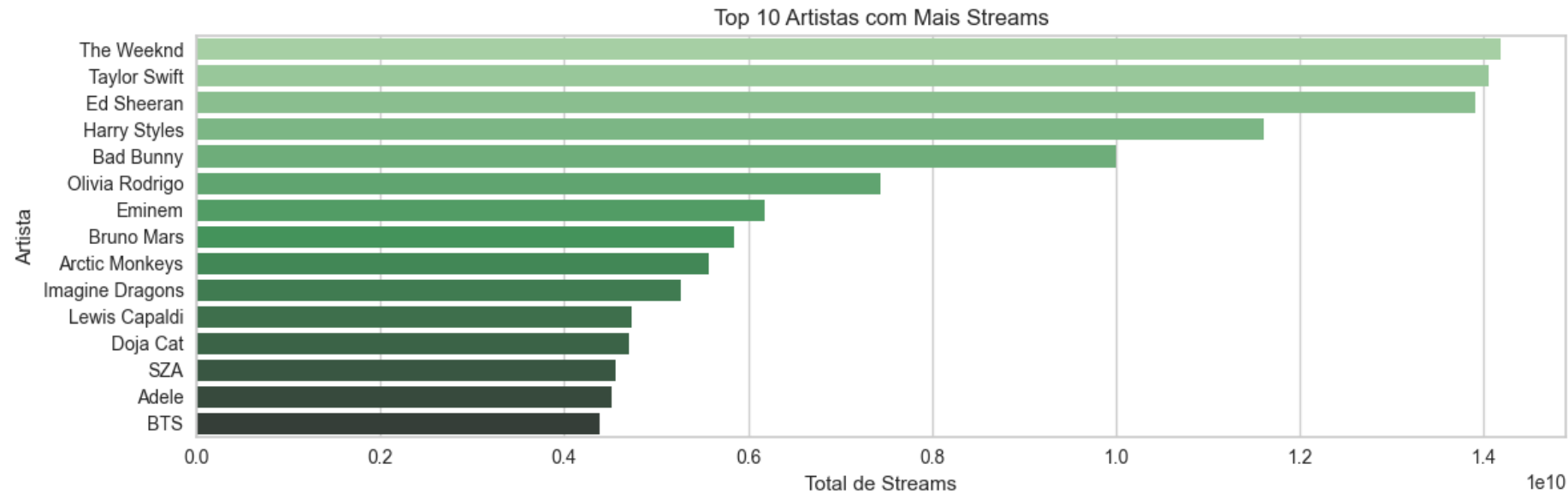
	artist_count	released_year	released_month	released_day	spotify_playlist	spotify_chart	apple_playlist	apple_chart	deezer_chart	bpm	danceability_%	valence_%	energy_%	acousticness_%	instrumentalness	liveness_%	speechiness_%
count	953.0	953.0	953.0	953.0	953.0	953.0	953.0	953.0	953.0	953.0	953.0	953.0	953.0	953.0	953.0	953.0	953.0
mean	1.56	2018.24	6.03	13.93	5200.12	12.01	67.81	51.91	2.67	122.54	66.97	51.43	64.28	27.06	1.58	18.21	10.13
std	0.89	11.12	3.57	9.2	7897.61	19.58	86.44	50.63	6.04	28.06	14.63	23.48	16.55	26.0	8.41	13.71	9.91
min	1.0	1930.0	1.0	1.0	31.0	0.0	0.0	0.0	0.0	65.0	23.0	4.0	9.0	0.0	0.0	3.0	2.0
25%	1.0	2020.0	3.0	6.0	875.0	0.0	13.0	7.0	0.0	100.0	57.0	32.0	53.0	6.0	0.0	10.0	4.0
50%	1.0	2022.0	6.0	13.0	2224.0	3.0	34.0	38.0	0.0	121.0	69.0	51.0	66.0	18.0	0.0	12.0	6.0
75%	2.0	2022.0	9.0	22.0	5542.0	16.0	88.0	87.0	2.0	140.0	78.0	70.0	77.0	43.0	0.0	24.0	11.0
max	8.0	2023.0	12.0	31.0	52898.0	147.0	672.0	275.0	58.0	206.0	96.0	97.0	97.0	97.0	91.0	97.0	64.0

### Observações do cientista:

- bpm: ritmo em sua maioria moderado e animado com valor médio aproximado de 120 BPM
- danceability\_%: média acima de 60%, indicando maioria de músicas dançantes
- energy\_%: média acima de 60%, indicando ritmo forte
- valence\_%: média acima de 50% indicando emoções equilibradas
- acousticness\_%: média baixa indicada pela alta quantidade de zeros, maioria das músicas produzidas por instrumentos eletrônicos
- instrumentalness\_%: média próxima de 2%, a maioria das músicas possui vocais
- speechiness\_%: média de 10% com máximo de 64% indicando presença de raps
- liveness\_%: média de 18% com 50%-75% das músicas com nível de liveness\_% médio variando entre 10%-24% com um máximo de 97% indicando a presença de músicas gravadas em ambientes abertos como shows e com efeitos de plateia

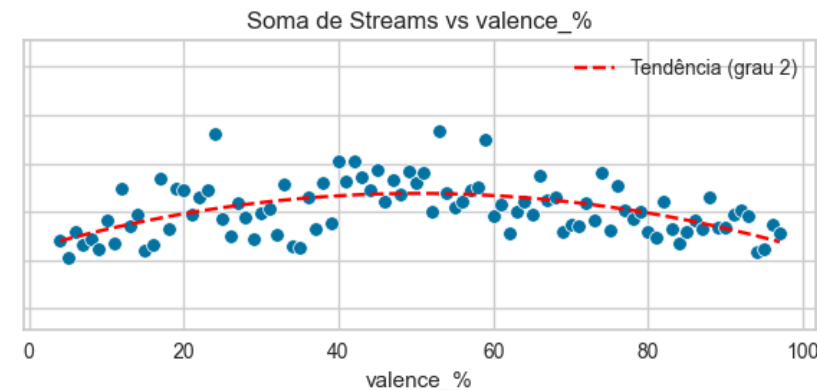
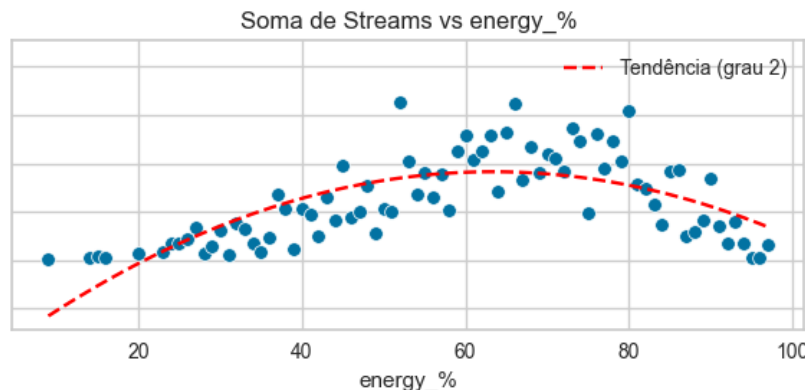
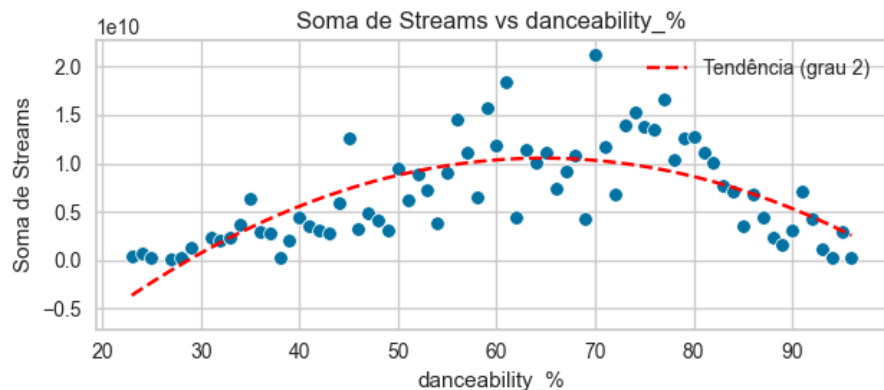


## Variáveis Categóricas





## Variáveis Numéricas



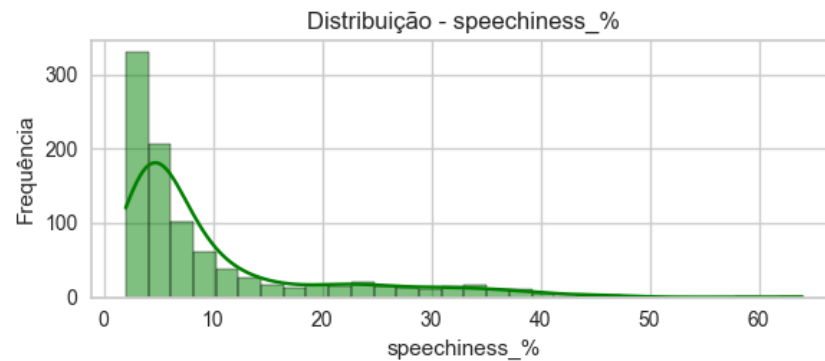
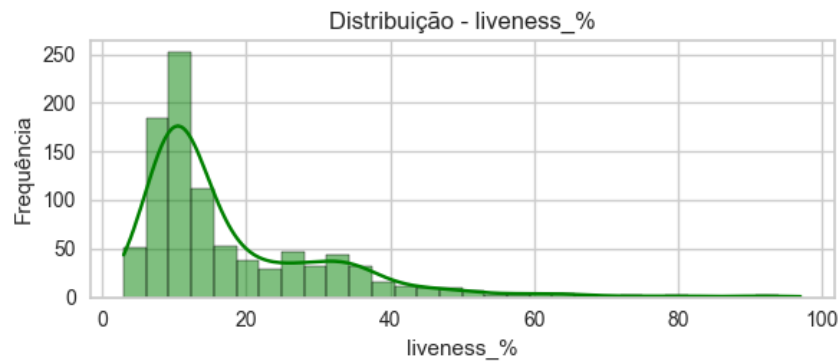
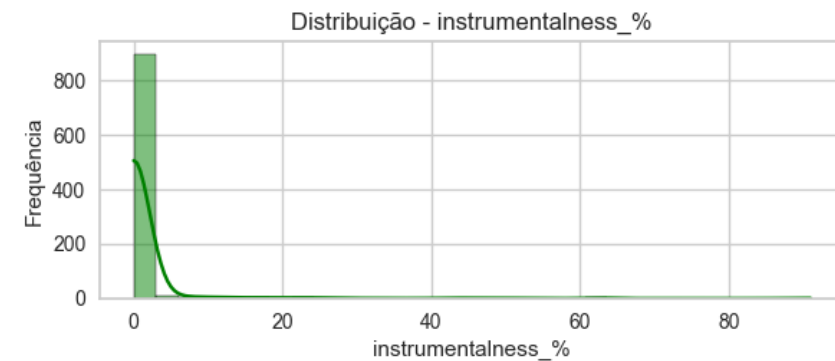
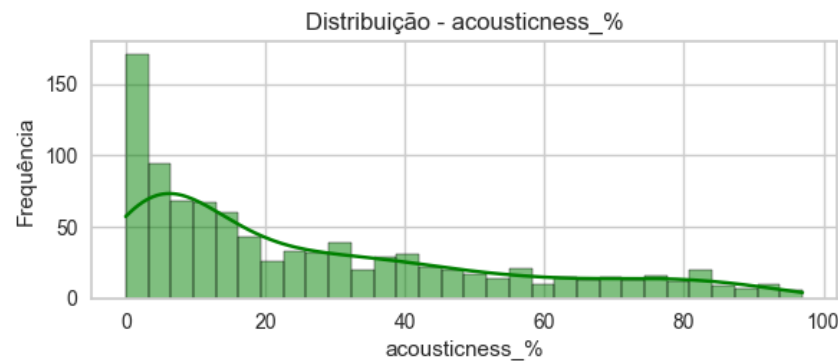
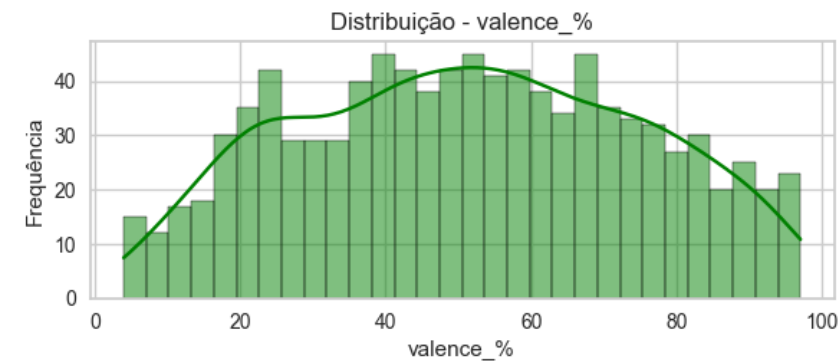
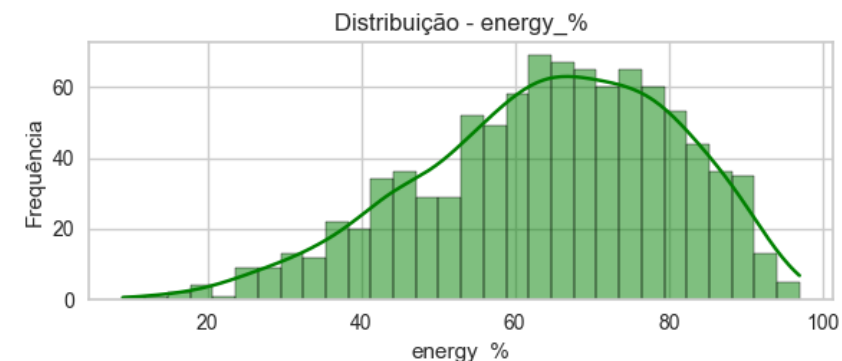
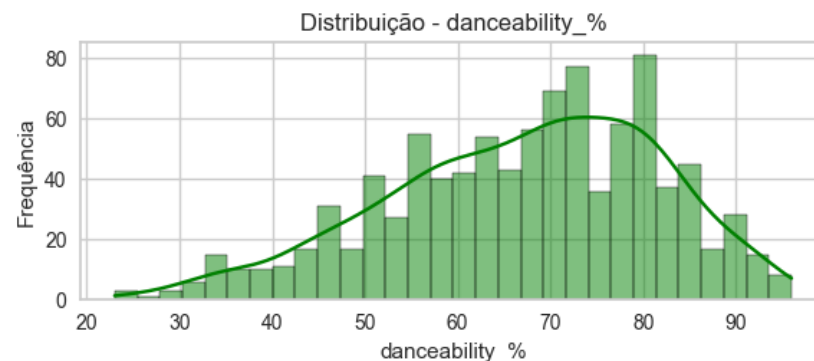
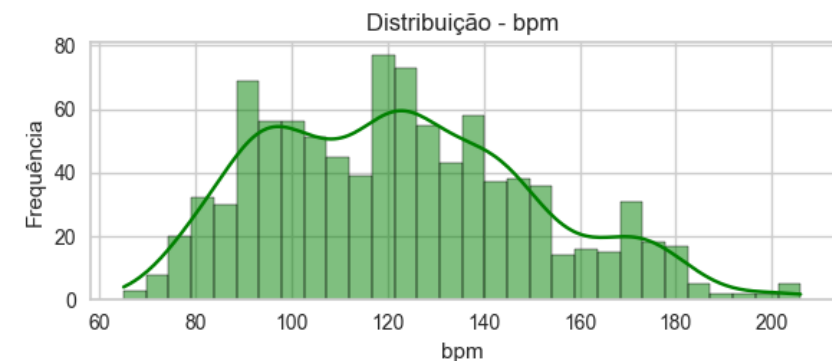
### Observações do cientista:

- Temos três curvas de tendência para danceability\_% e energy\_% com picos nos 60% e valence\_% temos o pico entre 40% a 60%, provavelmente as músicas mais populares ficam nesses intervalos de qualidade.

# Metodologia – Análise exploratória



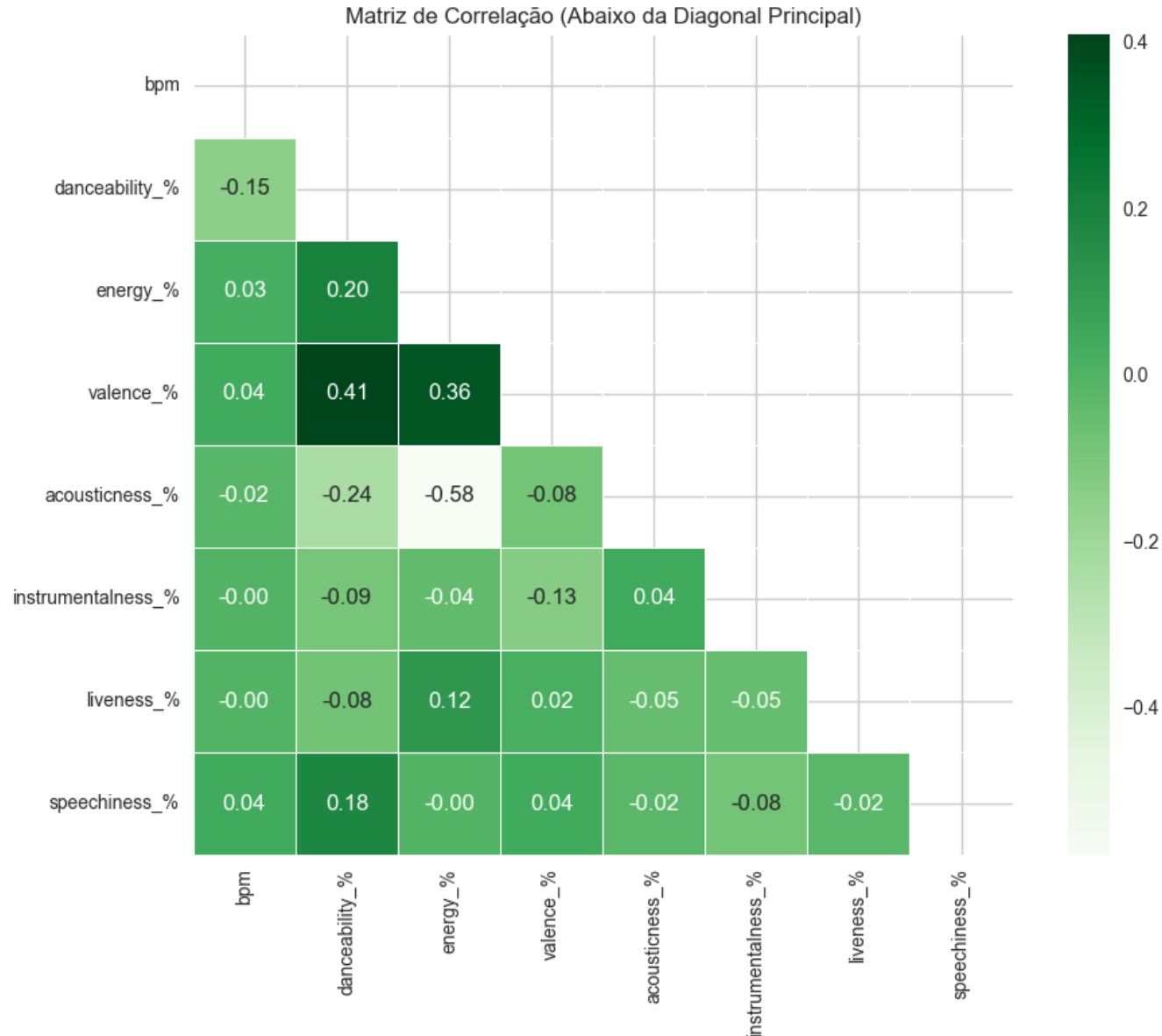
## Variáveis Numéricas – Distribuição







## Matriz de Correlação



### Observações do cientista:

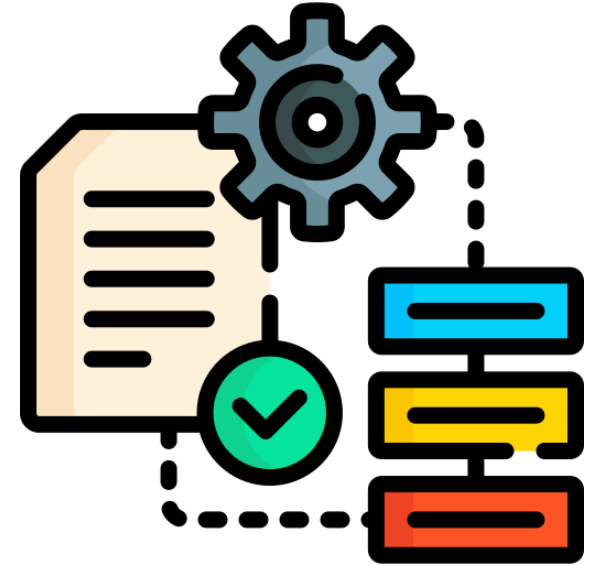
- Como energy\_% e acousticness\_% possuem correlação de -60% temos uma correlação inversa forte, podemos descartar do nosso modelo uma das duas variáveis.
- Como energy\_% de acordo com os gráficos da distribuição normal e boxplot possuem uma distribuição melhor, facilitando para o modelo durante a previsão.

# Metodologia – Pré-Processamento



Foi verificado e executado:

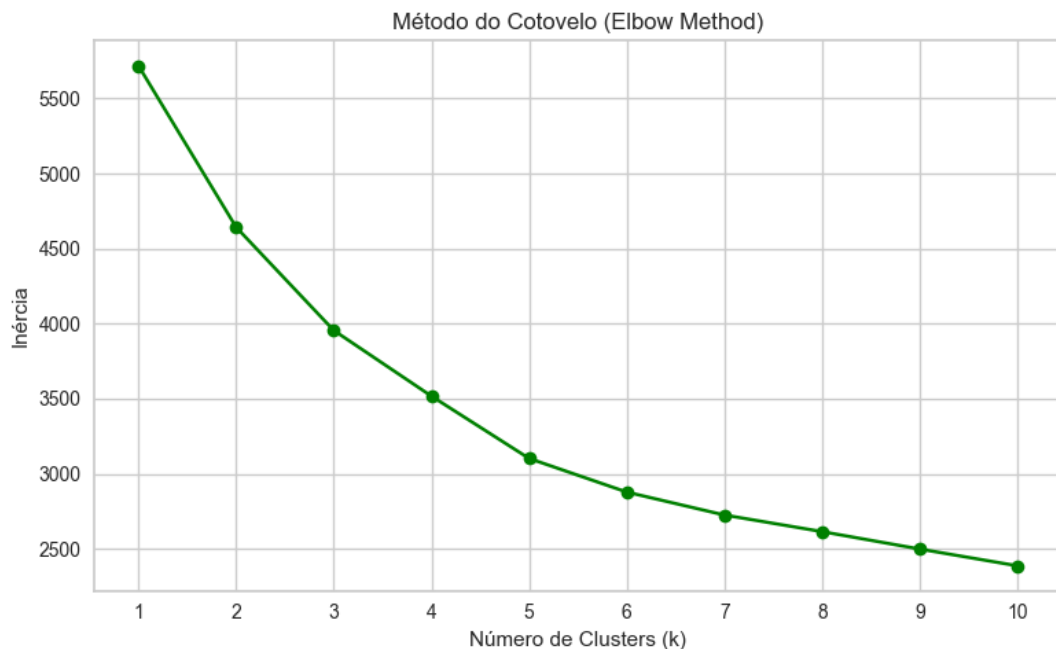
1. Verificação de nulos: in\_shazam\_charts e key
2. Valores nulos de key foram trocados por unknown
3. Valores nulos de in\_shazam\_charts foram trocados por 0 (nenhuma lista)
4. Todos os valores de colunas respectivamente numéricas foram convertidas para números, o mesmo para colunas categóricas



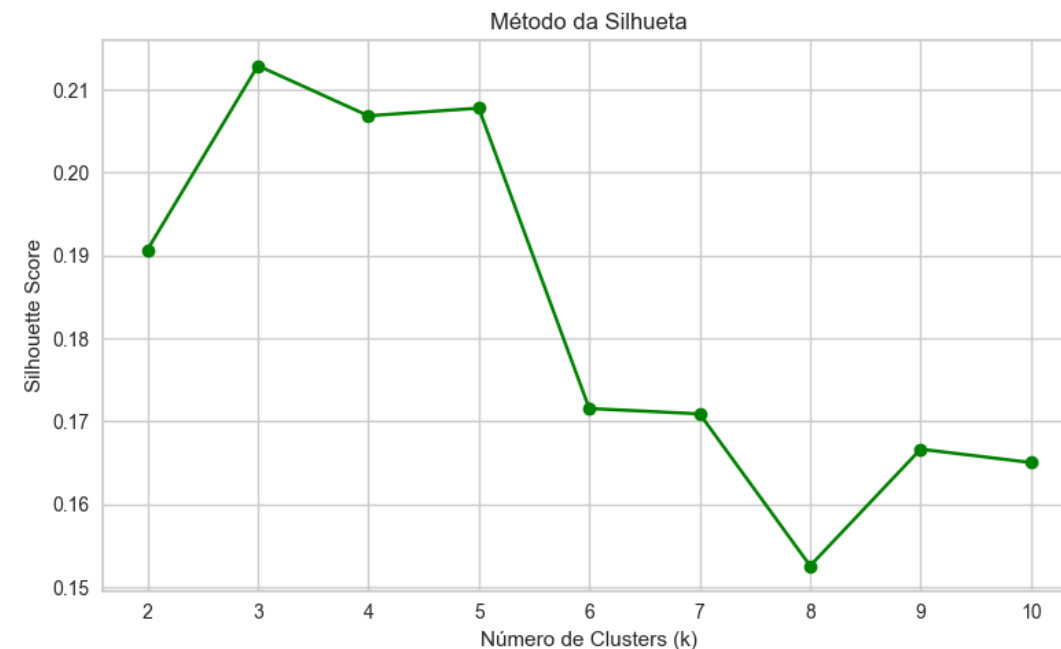
# Metodologia – Quantidade de Clusters



- Elbow



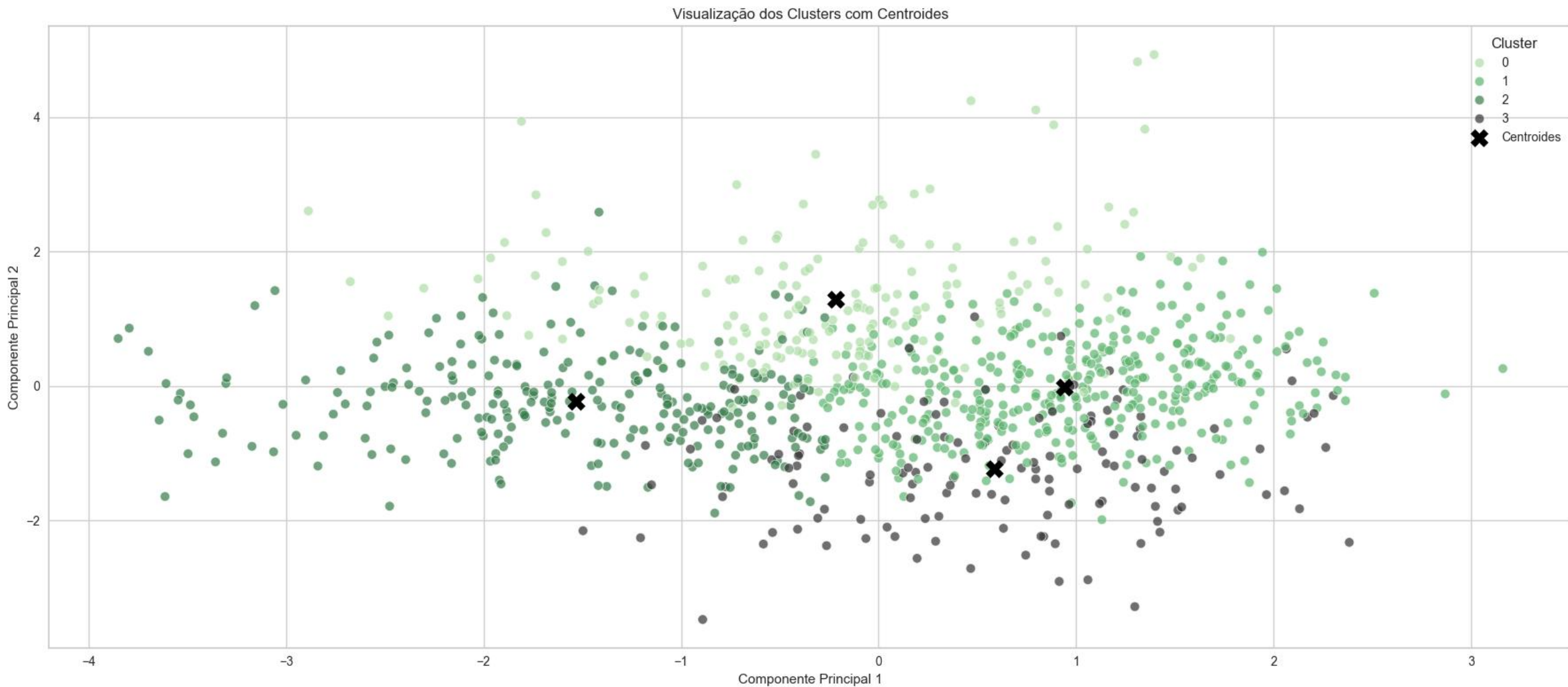
- Silhouette



## Observações do cientista:

- Pelos métodos seriam 3 clusters, no entanto a unidade de negócio pediu no mínimo 4. Portanto criaremos 4 cluster.

# Metodologia – Resultados do modelo



# Metodologia – Exemplos de cada Cluster



	track_name	artist(s)_name	cluster
0	Would've, Could've, Should've	Taylor Swift	0
1	LADY GAGA	Gabito Ballesteros, Junior H, Peso Pluma	0
2	Like Crazy	Jimin	0
3	Cruel Summer	Taylor Swift	0
4	WHERE SHE GOES	Bad Bunny	0
5	As It Was	Harry Styles	0
6	Celestial	Ed Sheeran	0
7	Typa Girl	BLACKPINK	0
8	El Azul	Junior H, Peso Pluma	0
9	I'm Good (Blue)	Bebe Rexha, David Guetta	0

	track_name	artist(s)_name	cluster
181	Flowers	Miley Cyrus	1
182	Super Shy	NewJeans	1
183	un x100to	Bad Bunny, Grupo Frontera	1
184	fukumean	Gunna	1
185	Ella Baila Sola	Eslabon Armado, Peso Pluma	1
186	You Right	Doja Cat, The Weeknd	1
187	Butter	BTS	1
188	Sad Girlz Luv Money Remix (feat. Kali Uchis)	Kali Uchis, Amaarae, Moliy	1
189	Life Goes On	Oliver Tree	1
190	Better Days (NEIKED x Mae Muller x Polo G)	NEIKED, Mae Muller, Polo G	1

	track_name	artist(s)_name	cluster
559	Love Nwantiti - Remix	Ckay, AX'EL, Dj Yo!	2
560	Lost in the Fire	The Weeknd, Gesaffelstein	2
561	Phantom Regret by Jim	The Weeknd	2
562	Dos Oruguitas	Sebastian Yatra	2
563	Daddy Issues	The Neighbourhood	2
564	Meet Me At Our Spot	THE ANXIETY, Willow, Tyler Cole	2
565	The Astronaut	Jin	2
566	My Mind & Me	Selena Gomez	2
567	we fell in love in october	girl in red	2
568	Miss You	Southstar	2

	track_name	artist(s)_name	cluster
819	Sprinter	Dave, Central Cee	3
820	TQG	Karol G, Shakira	3
821	TO THE MOON	Jnr Choi	3
822	Knife Talk (with 21 Savage ft. Project Pat)	Drake, Project Pat, 21 Savage	3
823	The Family Madrigal	Olga Merediz, Stephanie Beatriz, Encanto - Cast	3
824	Every Angel is Terrifying	The Weeknd	3
825	Privileged Rappers	Drake, 21 Savage	3
826	BackOutsideBoyz	Drake	3
827	Broke Boys	Drake, 21 Savage	3
828	California Breeze	Lil Baby	3

# Metodologia – Nomeando cada Cluster



## Observações do cientista:

	track_name	artist(s)_name	cluster	playlist_category
0	Seven (feat. Latto) (Explicit Ver.)	Latto, Jung Kook	1	Calmamente Pop
1	LALA	Myke Towers	1	Calmamente Pop
2	vampire	Olivia Rodrigo	2	Entre razões e emoções
3	Cruel Summer	Taylor Swift	0	Trilha das Estrelas
4	WHERE SHE GOES	Bad Bunny	0	Trilha das Estrelas
5	Sprinter	Dave, Central Cee	3	Flow Pesadão
6	Ella Baila Sola	Eslabon Armado, Peso Pluma	1	Calmamente Pop
7	Columbia	Quevedo	2	Entre razões e emoções
8	fukumean	Gunna	1	Calmamente Pop
9	La Bebe - Remix	Peso Pluma, Yng Lvcas	3	Flow Pesadão

### Nome de cada Cluster

- Trilha das Estrelas – Cluster 0
- Calmamente Pop – Cluster 1
- Entre razões e emoções – Cluster 2
- Flow Pesadão – Cluster 3



Obrigado!!!