

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Bruno Vieira

MONOGRAFIA EM SISTEMAS DE INFORMAÇÃO II

**QUAIS CARACTERÍSTICAS SÃO MAIS RELEVANTES PARA A POPULARIDADE
DE UMA MÚSICA EM DIFERENTES PAÍSES**

Belo Horizonte
2020

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação

**QUAIS CARACTERÍSTICAS SÃO MAIS RELEVANTES PARA A
POPULARIDADE DE UMA MÚSICA EM DIFERENTES PAÍSES**

por

Bruno Vieira

Monografia em Sistemas de Informação II

Apresentado como requisito da disciplina de Monografia em
Sistemas de Informação II do Curso de Bacharelado em Sistemas
de Informação da UFMG

Prof. Dr. Flavio Vinicius Diniz de Figueiredo
Orientador

Belo Horizonte
2020

Aos meus pais, Carlos e Sandra,
que sempre me incentivaram a ir mais longe.
Ao meu namorado, Jean,
que foi paciente e me apoiou nos momentos que precisei.
E à todos os amigos que fiz na UFMG,
que foram fundamentais para que eu chegasse até aqui.

RESUMO

Neste trabalho, são realizadas análises quantitativas de dados referentes aos atributos musicais *acousticness*, *danceability*, *energy*, *instrumentalness*, *liveness*, *speechiness* e *valence*, fornecidos por meio da API web do Spotify para desenvolvedores, com o objetivo de verificar a relevância de cada um desses atributos para a popularidade de uma música em países diferentes. Como fonte para as análises desta pesquisa, é construído um dataset baseado em dados disponíveis no Kaggle. Essa fonte de dados contém uma lista com mais de 3 milhões de músicas presentes nas paradas musicais top 200 de 53 países, coletadas ao longo de duas semanas no ano de 2017, juntamente com seus atributos musicais. A partir disso, é realizada uma análise exploratória utilizando bibliotecas do Python, para entender melhor os padrões gerais dos dados, seguido da aplicação do algoritmo de aprendizado de máquina k-means, para verificar a afinidade entre os países em relação aos seus atributos. Como resultado, foi observada grande relevância dos atributos musicais *danceability* e *energy* para todos os países analisados, além de indícios de atributos que são mais presentes em alguns países do que em outros.

Palavras-chave: Análise de dados. Música. Aprendizado não supervisionado. Características musicais. Spotify.

ABSTRACT

In this paper, quantitative data analysis are performed based on the musical features acousticness, danceability, energy, instrumentalness, liveness, speechiness and valence, provided by the Spotify web API for developers, in order to verify the relevance of each of these features for the popularity of a song in different countries. As a datasource for the analysis of this research, a dataset is built based on data available in the website Kaggle. This data source contains a list of more than 3 million songs on the top 200 music charts in 53 countries, collected over two weeks in 2017, along with their musical attributes. From this, an exploratory analysis is performed using Python libraries, to better understand the general patterns of the data, followed by the application of the k-means machine learning algorithm, to verify the affinity between countries in relation to their attributes. As a result, great relevance of the musical attributes danceability and energy was observed for all the countries analyzed, in addition to indications of attributes that are more present in some countries than in others.

Keywords: Data analysis. Music. Unsupervised learning. Audio features. Spotify.

LISTA DE FIGURAS

Figura 1 - Visão resumida da tabela de dados gerada.....	13
Figura 2 - Boxplot da distribuição dos atributos musicais no Spotify Global.....	13
Figura 3 - Média dos atributos musicais no Spotify Global.....	14
Figura 4 - Valor de K x Valor da Inertia.....	15
Figura 5 - Valor de K x Diferenças entre K e K-1.....	15
Figura 6 - Boxplot da distribuição dos atributos musicais no Spotify do Brasil.....	17
Figura 7 - Boxplot da distribuição dos atributos musicais no Spotify da Colômbia....	17
Figura 8 - Média entre países do atributo musical liveness.....	18
Figura 9 - Média entre países do atributo musical valence.....	19

LISTA DE TABELAS

Tabela 1 - Maiores e menores médias por atributo musical.....	20
Tabela 2 - P-valores obtidos no teste do qui quadrado dos 7 grupos.....	20
Tabela 3 - Grupo g2.....	21
Tabela 4 - Grupo g4.....	21
Tabela 5 - Grupo g6.....	22

SUMÁRIO

RESUMO.....	4
ABSTRACT.....	5
LISTA DE FIGURAS.....	6
LISTA DE TABELAS.....	7
1 INTRODUÇÃO.....	9
2 TRABALHOS RELACIONADOS.....	10
3 METODOLOGIA.....	11
3.1 Coleta dos Dados.....	11
3.2 Consolidação dos Dados.....	12
3.3 Análise Exploratória.....	13
3.4 Agrupamento com K-means.....	15
4 RESULTADOS E DISCUSSÕES.....	16
4.1 Análise Exploratória.....	16
4.2 Agrupamento com K-means.....	20
5 CONCLUSÃO.....	22
6 REFERÊNCIAS.....	23
APÊNDICE.....	25

1 INTRODUÇÃO

O Spotify é um serviço de streaming de música e, mais recentemente, podcasts, que dá acesso à milhões de conteúdos por meio de um plano gratuito ou por opções pagas. Dentro dessa grande quantidade de conteúdo disponível, a plataforma disponibiliza meios de sugerir músicas mais relevantes para seus usuários utilizando sistemas de recomendação e também fornecendo *playlists* com as paradas de sucesso atuais.

Conhecer os critérios e atributos que fazem uma música se tornar mais popular é potencialmente interessante, como mencionado por Al-Beitawi et al., já que este conhecimento pode fomentar melhorias na experiência e na satisfação do usuário final, garantindo melhores recomendações. O próprio Spotify disponibiliza por meio de uma API (Interface de Programação de Aplicações) alguns atributos referentes à cada música, caracterizando-as nos quesitos: *acousticness*, *danceability*, *energy*, *instrumentalness*, *liveness*, *loudness*, *speechiness*, *valence* e *tempo*. Estes atributos, apesar de representarem qualitativamente características de cada música, são expressos em escalas quantitativas que têm limites superiores e inferiores variados.

Para este estudo, especificamente, são focados os sete atributos que possuem variação numérica entre 0 e 1, devido à maior facilidade de análise e comparação entre eles. São estes: (I) *acousticness*, dado como uma medida de confiança do quão acústica é uma canção; (II) *danceability*, que descreve o quão dançante uma música é levando em consideração elementos como tempo, estabilidade do ritmo e força da batida; (III) *energy*, que é a percepção da intensidade de uma música, dada por fatores como altura, timbre e gama dinâmica; (IV) *instrumentalness*, que prevê o uso maior ou menor de vocais em um canção, (V) *liveness*, que representa a probabilidade de uma música ter sido tocada ao vivo, dada a presença de vozes de uma plateia na gravação; (VI) *speechiness*, que detecta a quantidade de palavras faladas em uma música; e, por fim, (VII) *valence*, que representa a positividade transmitida por uma faixa musical.

A partir disso, são feitas análises quantitativas sobre esses atributos em uma base de dados musicais retirada do site Kaggle, visando verificar padrões entre diferentes países. Esses padrões são verificados em duas etapas: Na primeira, uma

análise exploratória é feita e exemplificada com diferentes visualizações de dados, como boxplots e histogramas; Na segunda etapa, os dados são analisados, por meio de um agrupamento gerado pelo algoritmo k-means, e a qualidade dessa análise é mensurada e discutida, no intuito de responder a questão inicial desta pesquisa.

2 TRABALHOS RELACIONADOS

Alguns trabalhos sobre a popularidade de músicas no Spotify e a análise de seus atributos já foram conduzidos, seguindo diferentes perspectivas. No trabalho de Suh, por exemplo, foram utilizados os atributos quantitativos já mencionados anteriormente (*acousticness*, *danceability*, *energy*, *instrumentalness*, *liveness*, *loudness*, *speechiness*, *valence* e *tempo*) para determinar o sucesso de faixas na parada musical das 200 melhores músicas do Spotify, utilizando regressões e levando em consideração as divergências em cinco países diferentes, tendo como fator de popularidade a estabilidade da faixa em uma determinada posição na parada musical.

Já no trabalho de Al-Beitawi et al., apesar do uso dos mesmos atributos quantitativos, a abordagem escolhida para análise foi a separação das faixas em clusters e a utilização do algoritmo K-means. Os resultados mostraram que os clusters que possuem músicas com o quesito *danceability* mais alto tendem a estar no topo de paradas musicais.

Um dos papéis deste trabalho é analisar as diferenças culturais na esfera musical em diferentes países e, com isso, é relevante entender como a música popular evolui ao longo do tempo e se transforma, assim como a cultura de um povo no geral. Os trabalhos de Mauch et al. e de Vigliensoni e Fujinaga investigam essa evolução por meio de uma análise da parada musical Billboard Hot 100 entre os anos de 1960 e 2010 e pela criação de um dataset unificado utilizando dados extraídos do Last.fm, respectivamente. Em ambos os trabalhos temos um retrato da realidade musical, sendo o primeiro mais focado nos pontos-chaves de mudanças, como as "revoluções" do pop ocorridas em 1964, 1983 e 1991 e o segundo focando em características mais demográficas, produzindo um dataset útil para eventuais análises.

Assim como este trabalho, três estudos levaram em consideração a relação entre o gosto musical e a cultura de um país. Schedl et al. utilizaram a correlação de Spearman e QAP (Quadratic Assignment procedure) para medir a correlação dos fatores socioeconômicos de um país com o seu gosto musical, gerando resultados que podem melhorar sistemas de recomendação baseados em cultura. O artigo de Liu et al. seguiu o mesmo ideal, gerando indicadores curiosos, como a falta de relação significativa entre fatores geoeconômicos e as preferências por artistas, álbuns e gêneros. Por fim, o artigo de Zangerle utiliza as mesmas características musicais já mencionadas neste texto em conjunto com fatores culturais e socioeconômicos para aumentar a acurácia de um algoritmo de recomendação também baseado em cultura. O modelo foi desenvolvido usando dados sobre reprodução de música de cerca de 55 mil usuários de 36 países e, em seguida, foi aplicado em um sistema de recomendação já existente.

Em conclusão, os artigos de Silva, Thiago H. et al e Garcia-Gavilanes, apesar de não analisarem dados sobre preferências musicais, identificam outros padrões humanos relacionados à cultura e fatores socioeconômicos. Os algoritmos utilizados para mensuração de hábitos alimentares, assim como, padrões de comportamento no Twitter são úteis para influenciar este trabalho, além de nortear formas de lidar com os aspectos culturais presentes na pesquisa.

3 METODOLOGIA

3.1 Coleta dos Dados

Para a realização das análises deste trabalho, foi construído um dataset baseado em outra fonte de dados disponível no Kaggle. Essa fonte de dados contém uma lista com mais de 3 milhões de músicas presentes nas paradas musicais top 200 de 53 países, coletadas ao longo de duas semanas no ano de 2017. Resumidamente, os dados são compostos por 6629 artistas e 18598 canções únicas com informações sobre a posição na parada musical da faixa, seu número de

execuções, data de coleta do dado, região de origem e link para a plataforma do Spotify.

A partir disso, foram obtidos os atributos musicais para cada linha da base de dados do Kaggle, utilizando o Spotify Web API. Com o auxílio da biblioteca Spotipy e desta API, da própria plataforma, foram feitas requisições com os links únicos de cada canção e, os resultados, com os atributos de interesse, foram armazenados, em memória, em um notebook Jupyter, sendo posteriormente exportados para um arquivo CSV, com o intuito de facilitar o uso dos dados em longo prazo. Além disso, devido à limitação de uso do Spotify Web API, as requisições tiveram que ser separadas em lotes, sendo executadas em sequência e consolidadas em uma única matriz ao final do processo.

3.2 Consolidação dos Dados

Para o uso eficaz dos dados, foram feitos alguns tratamentos para padronizá-los. Inicialmente, a base de dados disponível no Kaggle, contendo as faixas musicais, foi unida com o arquivo CSV que possuía os atributos para cada uma dessas faixas, utilizando a biblioteca Pandas para o Python. Em seguida, algumas colunas que possuíam dados não relevantes para a pesquisa foram removidas, juntamente com as possíveis linhas que possuíam campos nulos, causados pela junção das tabelas, além de três faixas específicas, que não possuíam informações referentes aos seus atributos na API Web do Spotify. São elas: (I) One Love Manchester (Live from One Love Manchester); (II) Año Nuevo 2017; e (III) Año Nuevo 2018.

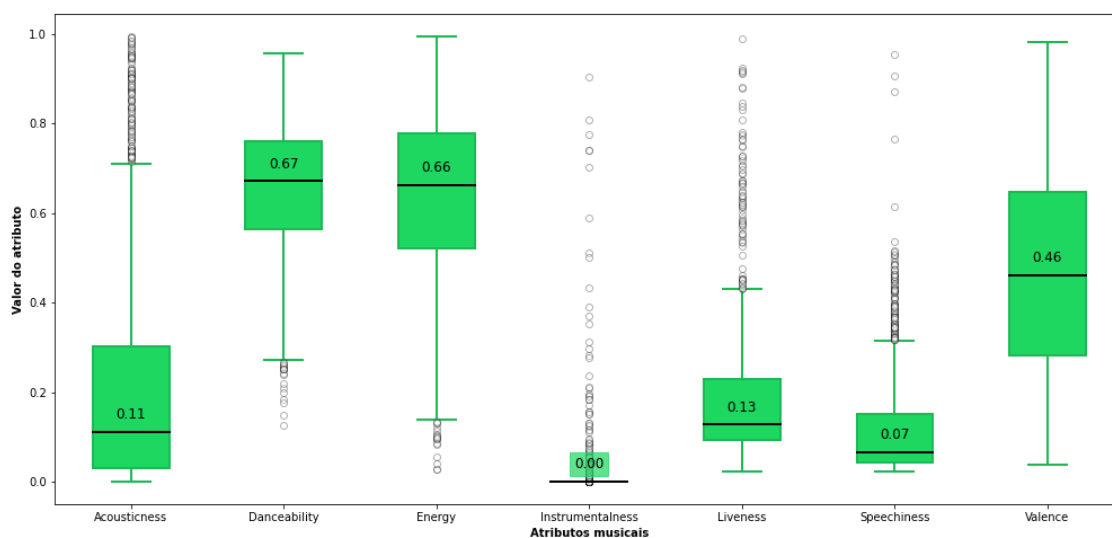
Por fim, foi realizada uma última junção entre tabelas para inserir uma nova coluna com o nome do país referente à faixa musical, visto que, existia apenas a informação do código de cada país, não sendo fácil identificar visualmente a qual localidade cada código se refere. Como resultado, uma tabela consolidada com 3.440.521 linhas e 21 colunas foi gerada, pronta para ser consultada e manipulada conforme os interesses da pesquisa, como é mostrado na figura 1.

Figura 1 - Visão resumida da tabela de dados gerada

Position	Track Name	Artist	Streams	Date	Region	Country	danceability	energy	key	...	mode	speechiness	acousticness	instrumentalness	liveness
1	Reggaetón Lento (Bailamos)	CNCO	19272	2017-01-01	ec	Ecuador	0.761	0.838	4.0	...	0.0	0.0502	0.40000	0.000000	0.176
2	Chantaje	Shakira	19270	2017-01-01	ec	Ecuador	0.852	0.773	8.0	...	0.0	0.0776	0.18700	0.000030	0.159
3	Otra Vez (feat. J Balvin)	Zion & Lennox	15761	2017-01-01	ec	Ecuador	0.832	0.772	10.0	...	1.0	0.1000	0.05590	0.000486	0.440
4	Vente Pa' Ca	Ricky Martin	14954	2017-01-01	ec	Ecuador	0.663	0.920	11.0	...	0.0	0.2260	0.00431	0.000017	0.101
5	Safari	J Balvin	14269	2017-01-01	ec	Ecuador	0.508	0.687	0.0	...	1.0	0.3260	0.55100	0.000003	0.126

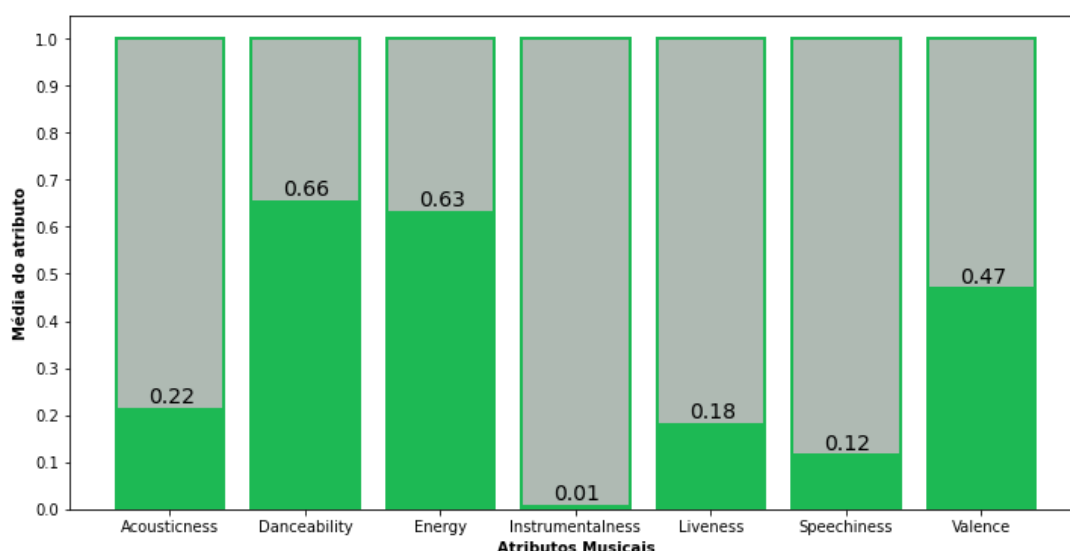
3.3 Análise Exploratória

Em princípio, foi realizada uma análise exploratória para identificar o padrão de comportamento dos dados, a fim de entender melhor como estes se distribuem e se já existiam tendências que relacionam as músicas aos respectivos países. Para esta e posteriores análises foram selecionados apenas os atributos musicais *acousticness*, *danceability*, *energy*, *instrumentalness*, *liveness*, *speechiness* e *valence*, visto que, estes atributos possuem uma escala que varia entre 0 e 1, diferente de outros disponibilizados pelo Spotify, facilitando a análise e comparação entre eles.

Figura 2 - Boxplot da distribuição dos atributos musicais no Spotify Global

Para esta análise, foram gerados boxplots, como ilustrado pela figura 2, que refletem a distribuição de cada uma das 7 features musicais, tanto em um cenário global, quanto em cenários individuais para cada país. Para montagem dos gráficos, foram filtradas as músicas para o país de interesse e selecionadas apenas as entradas únicas, sem repetir aquelas que aparecem mais de uma vez em diferentes datas na parada musical do país. De forma complementar, outro gráfico evidenciando a tendência central de média dos dados foi gerado, como mostra a figura 3.

Figura 3 - Média dos atributos musicais no Spotify Global



Por fim, com a intenção de verificar, de forma mais aprofundada, as possíveis peculiaridades em cada país, a base de dados foi filtrada, mais uma vez, de modo a selecionar, para cada país, apenas as músicas únicas que não estão presentes nas paradas musicais de nenhum outro local no mundo. A partir dessa filtragem, foram elaborados novos boxplots com pequenas diferenças que serão discutidas posteriormente.

3.4 Agrupamento com K-means

Nesta segunda etapa, o objetivo estabelecido foi verificar a aderência dos dados, por meio da criação de grupos, utilizando o algoritmo de aprendizado não supervisionado K-means. Para realizar essa tarefa, foram selecionados os mesmos dados anteriores, contendo as músicas únicas de cada país e totalizando 15.080 entradas.

Para definir, empiricamente, o número ideal de grupos a serem gerados com o algoritmo, foram executadas 19 iterações, variando o valor de k entre 2 e 20. A partir disso, foi observado o valor de retorno da *inertia*, que indica a qualidade do grupo criado. Nos gráficos ilustrados nas figuras 4 e 5, podemos notar que, aproximadamente entre 6 e 8, o valor de k tende a estabilizar o algoritmo, não gerando grandes diferenças entre os valores de inertia. Dada essa conclusão, o valor ideal de K escolhido foi o 7.

Figura 4 - Valor de K x Valor da Inertia

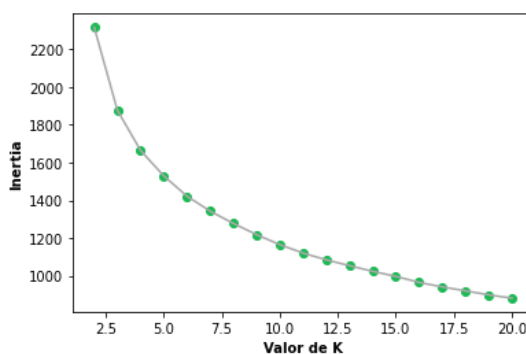
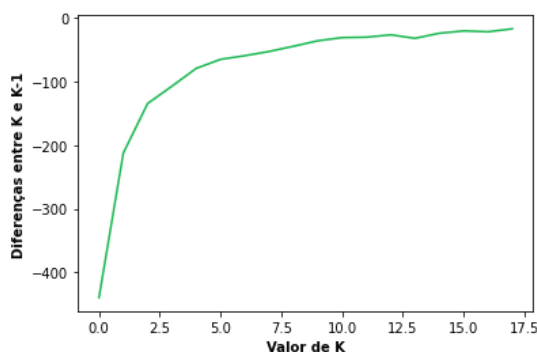


Figura 5 - Valor de K x Diferenças entre K e K-1



Em conclusão, para cada um dos 7 grupos gerados, foram criadas tabelas comparativas contendo os países pertencentes àquele grupo e a porcentagem de músicas referentes a cada país. Desta maneira, é possível verificar a aderência de cada localidade ao grupo estabelecido. Além disso, para assegurar que os dados de cada grupo tendem a seguir a mesma frequência dos dados do dataset original, foram executados 7 testes do qui quadrado, que têm seus resultados discutidos posteriormente.

4 RESULTADOS E DISCUSSÕES

4.1 Análise Exploratória

A partir dos primeiros gráficos de boxplot gerados sobre os dados, pode-se perceber uma tendência geral entre as músicas ao redor do mundo. No gráfico da figura 2, por exemplo, são evidentes os valores de mediana mais baixos nos atributos *acousticness*, *instrumentalness*, *liveness* e *speechiness*, enquanto as medianas dos atributos *danceability* e *energy*, são mais elevadas, demonstrando que músicas mais animadas e com grande potencial para serem dançadas, têm maior relevância para serem consideradas populares, segundo as paradas musicais do Spotify. O atributo *valence* tende a se estabilizar em valores centrais entre 0.4 e 0.6, mostrando-se mais neutro em relação aos outros.

Contudo, como destacado anteriormente, os primeiros boxplots gerados evidenciam uma tendência geral dos dados, não sendo possível perceber grandes diferenças nos padrões de países distintos. Seguindo disso, foram elaborados novos boxplots, nos quais os dados de cada país possuem apenas músicas únicas, que não estão presentes nas paradas musicais de nenhum outro local. Novamente, a apresentação geral dos dados pareceu seguir o mesmo padrão, com algumas diferenças sutis, como pode ser percebido nas figuras 6 e 7. Na distribuição dos dados no Brasil, por exemplo, é possível perceber que o valor do atributo *liveness* entre a mediana e o terceiro quartil é maior do que em outros exemplos apresentados, dando indícios de que a popularidade de músicas ao vivo no país é

maior do que em outros lugares. Em adição, é possível notar um indicativo de que, no Spotify da Colômbia, as músicas tendem a ter um valor de *valence* muito superior a outras localidades, com cerca de 38 pontos percentuais a mais que o Spotify Global, por exemplo.

Figura 6 - Boxplot da distribuição dos atributos musicais no Spotify do Brasil

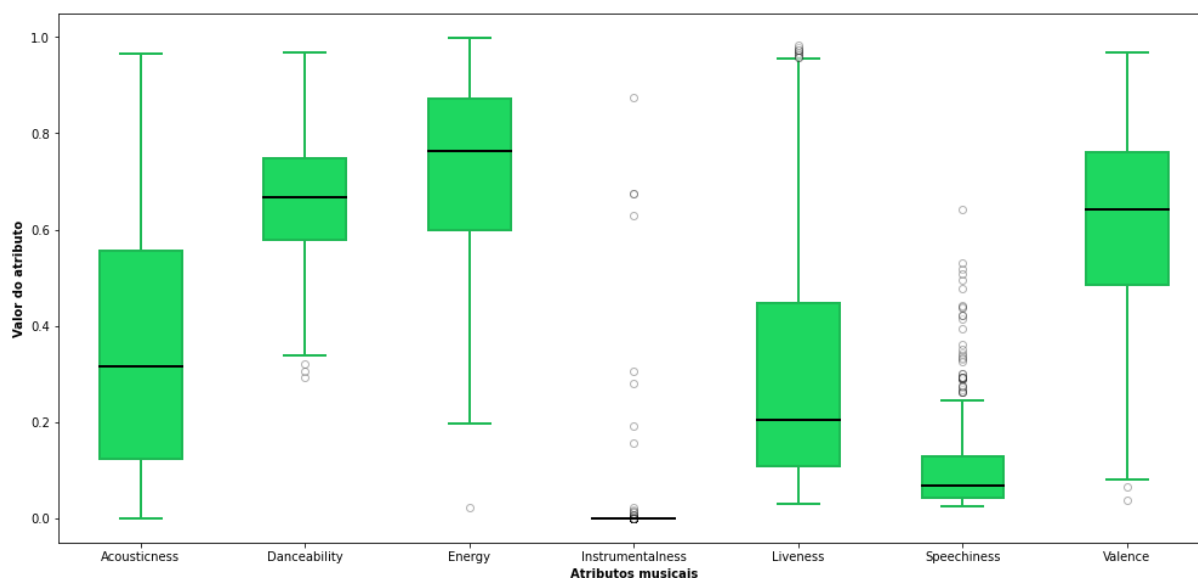
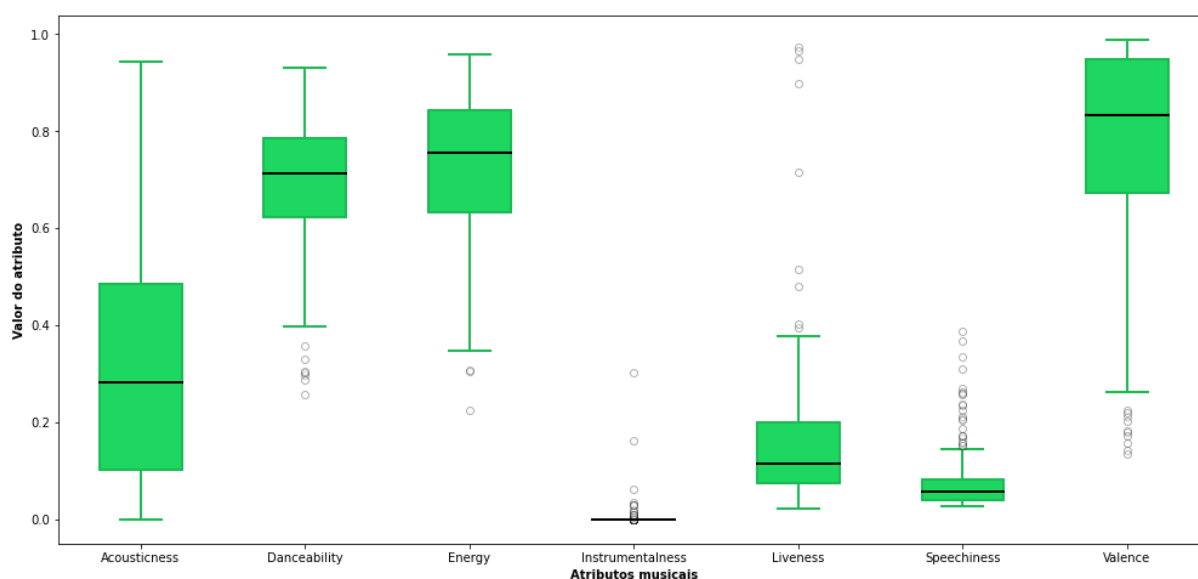


Figura 7 - Boxplot da distribuição dos atributos musicais no Spotify da Colômbia



Por fim, como suporte às evidências já apresentadas nos boxplots anteriores, as figuras 8 e 9 apresentam, respectivamente, a média para os valores dos atributos *liveness* e *valence*. É possível verificar que o Brasil, de fato, possui o índice de músicas executadas ao vivo mais elevado que outros países, posicionando-se, inclusive, no topo desta categoria. Para o atributo *valence*, também é possível confirmar o posicionamento elevado da Colômbia, como salientado anteriormente.

Figura 8 - Média entre países do atributo musical liveness

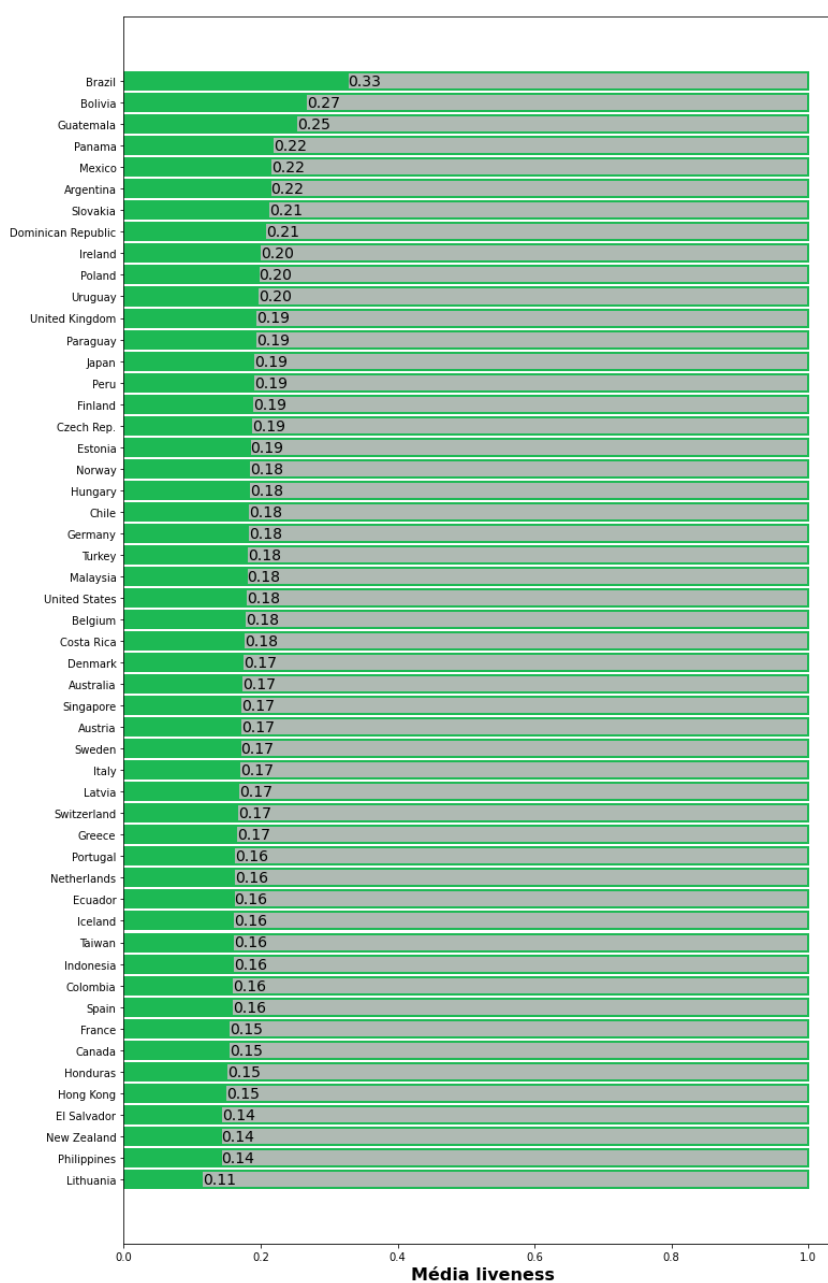
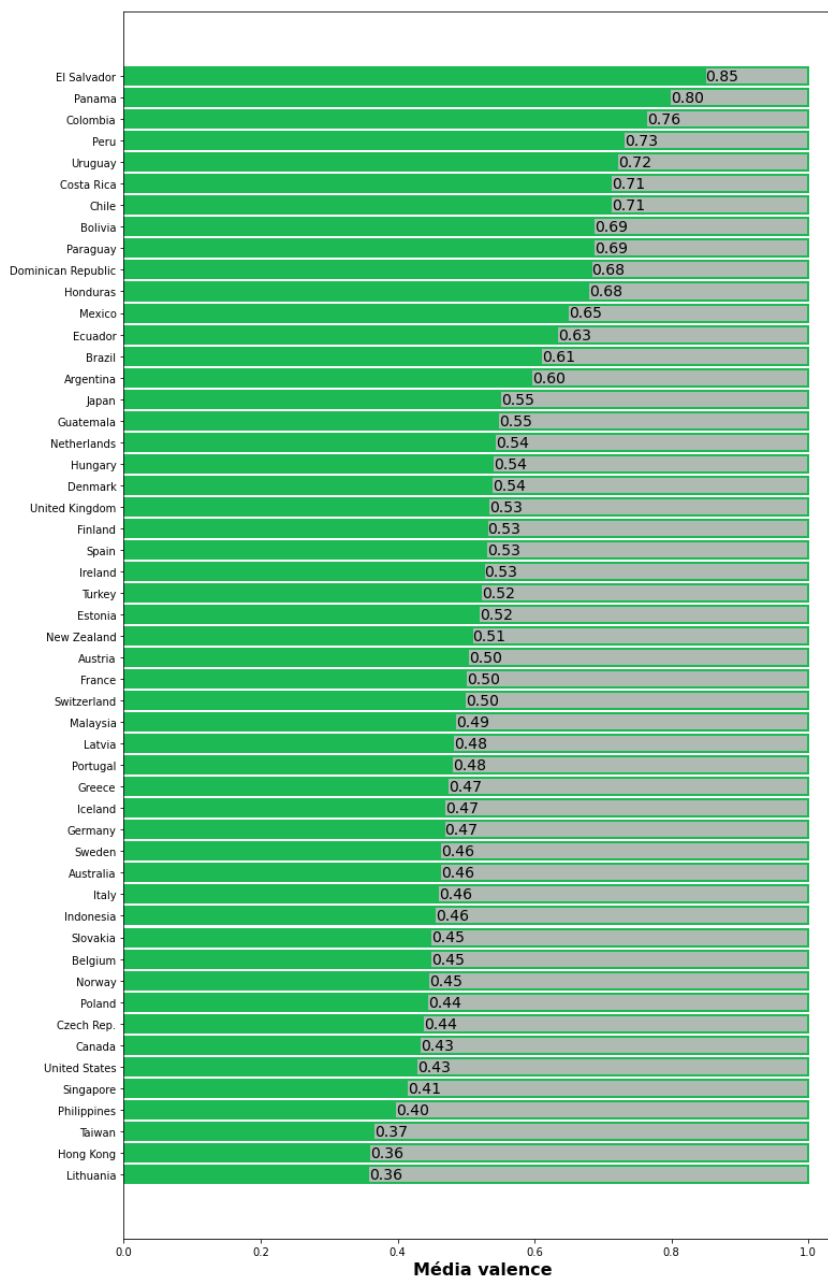


Figura 9 - Média entre países do atributo musical valence

Em resumo, na tabela 1, são apresentados os países com maior e menor índice, em média, de todos os 7 atributos analisados nesta pesquisa. Os dados mais detalhados para cada atributo podem ser acessados no repositório disponível no apêndice deste trabalho.

Tabela 1 - Maiores e menores médias por atributo musical

Atributo	País com menor índice (em média)	Valor do menor índice (em média)	País com maior índice (em média)	Valor do maior índice (em média)
Acousticness	Lituânia	0.11	Hong Kong	0.53
Danceability	Hong Kong	0.54	El Salvador	0.77
Energy	Hong Kong	0.48	Bolívia	0.77
Instrumentalness	Bolívia	0.00	Equador	0.10
Liveness	Lituânia	0.11	Brasil	0.33
Speechiness	Hong Kong	0.04	Alemanha	0.22
Valence	Lituânia	0.36	El Salvador	0.85

4.2 Agrupamento com K-means

A partir dos grupos criados com o algoritmo k-means, como mencionado na seção de metodologia desse texto, foram executados testes do qui quadrado, com a intenção de garantir que os dados agrupados seguem a mesma frequência de distribuição dos dados originais, como dita a hipótese nula do teste. Na tabela 2, a seguir, é possível verificar o p-valor para cada um dos 7 grupos testados, onde, valores mais próximos de 1 indicam melhores resultados, assegurando que os dados possuem baixa probabilidade de não seguir o padrão original.

Tabela 2 - P-valores obtidos no teste do qui quadrado dos 7 grupos

Grupo	g0	g1	g2	g3	g4	g5	g6
P-valor	0.87427	0.00224	0.95879	0.00066	0.99588	0.77290	0.99999

Como mostra a tabela 2, os grupos g1 e g3 possuem um p-valor muito próximo de 0, podendo ser facilmente descartados da análise. Dos grupos restantes, adotou-se uma heurística para considerar apenas aqueles com p-valor superior a 95%, fazendo com que os grupos g0 e g5 fossem, também, descartados. Por fim, os grupos resultantes dessa filtragem, g2, g4 e g6, são ilustradas pelas tabelas 3, 4 e 5 a seguir.

Tabela 3 - Grupo g2

País	Número de músicas	Porcentagem (número de músicas / total)
Taiwan	362	19.87%
Hong Kong	117	6.42%
França	99	5.43%
Itália	80	4.39%
Países baixos	79	4.34%
Outros	1085	59.55%
Total	1822 músicas	

Tabela 4 - Grupo g4

País	Número de músicas	Porcentagem (número de músicas / total)
Japão	278	11.42%
Taiwan	233	9.57%
Finlândia	232	9.53%
Itália	210	8.63%
Suécia	180	7.39%
Outros	1301	53.46%
Total	2434 músicas	

Tabela 5 - Grupo g6

País	Número de músicas	Porcentagem (número de músicas / total)
Finlândia	354	9.19%
Japão	305	7.92%
Países Baixos	281	7.29%
Suécia	205	5.32%
Itália	193	5.01%
Outros	2514	65.27%
Total	3852 músicas	

Pode-se notar que, Taiwan possui grande representatividade em g2 e g4, assim como a Finlândia está bem colocada em todos os 3 grupos apresentados. Isso mostra que os grupos não são tão específicos e que alguns países possuem tanta diversidade nas características musicais que não podem ser majoritariamente encaixados em um único agrupamento. Além disso, não há nenhuma tendência completamente evidente nos dados das tabelas. O grupo g2, por exemplo, tem grande representatividade de 2 países orientais em paralelo com 3 países europeus culturalmente distintos.

5 CONCLUSÃO

Com base nos resultados discutidos, foi possível obter alguns indícios para responder à questão levantada pelo título da pesquisa. Por meio da análise exploratória e gráficos da distribuição dos dados, foi possível perceber que, no geral, a "dançabilidade" e a "energia" são atributos relevantes para a popularidade de uma música em todos os países, sendo apresentados com valores bem elevados na média e mediana. Além disso, gráficos complementares mostraram que alguns

atributos são mais presentes em alguns países do que em outros, como evidenciado pela tabela 1.

Por fim, os grupos gerados pelo algoritmo k-means mostraram afinidades musicais entre países como Taiwan, Hong Kong e França, não elucidando uma correlação clara entre esses locais. Para estudar melhor essas relações, trabalhos futuros podem focar em outras técnicas de aprendizado supervisionado e não supervisionado, também sendo possível utilizar uma base de dados diferente, de modo a comparar os resultados obtidos com este trabalho.

6 REFERÊNCIAS

O que é o Spotify?. Disponível em:

<https://support.spotify.com/br/using_spotify/getting_started/what-is-spotify/> Acesso em: 19 Ago. 2020

SUH, Brendan Joseph. International Music Preferences: An Analysis of the Determinants of Song Popularity on Spotify for the US, Norway, Taiwan, Ecuador, and Costa Rica. 2019.

AL-BEITAWI, Zayd; SALEHAN, Mohammad; ZHANG, Sonya. Cluster Analysis of Musical Attributes for Top Trending Songs. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020.

ARAUJO, Carlos Soares; CRISTO, Marco; GIUSTI, Rafael. Predicting Music Popularity on Streaming Platforms. In: Anais do XVII Simpósio Brasileiro de Computação Musical. SBC, 2019. p. 141-148.

VIGLIENSONI, Gabriel; FUJINAGA, Ichiro. The Music Listening Histories Dataset. In: ISMIR. 2017. p. 96-102.

GARCIA-GAVILANES, R. O. G. G.; QUERCIA, Daniele; JAIMES, Alejandro. Cultural dimensions in twitter: Time, individualism and power. 2013.

SILVA, Thiago H. et al. You are what you eat (and drink): Identifying cultural boundaries by analyzing food & drink habits in foursquare. arXiv preprint arXiv:1404.1009, 2014.

SCHEDL, Markus et al. Indicators of country similarity in terms of music taste, cultural, and socio-economic factors. In: 2017 IEEE International Symposium on Multimedia (ISM). IEEE, 2017. p. 308-311.

LIU, Meijun; HU, Xiao; SCHEDL, Markus. The relation of culture, socio-economics, and friendship to music preferences: A large-scale, cross-country study. PloS one, v. 13, n. 12, p. e0208186, 2018.

ZANGERLE, Eva; PICHL, Martin; SCHEDL, Markus. User Models for Culture-Aware Music Recommendation: Fusing Acoustic and Cultural Cues. Transactions of the International Society for Music Information Retrieval, v. 3, n. 1, 2020.

MAUCH, Matthias et al. The evolution of popular music: USA 1960–2010. Royal Society open science, v. 2, n. 5, p. 150081, 2015.

APÊNDICE

Todo o trabalho e a análise feita sobre os dados podem ser acessados no seguinte repositório: <https://github.com/brunoa15/monografia-2>

A base de dados está disponível em:

<https://drive.google.com/drive/folders/1mKeOyff6iVSQiA3eiNcLONJyhvOupe2W?usp=sharing>