



**UEPB**

**UNIVERSIDADE ESTADUAL DA PARAÍBA  
CAMPUS I  
CENTRO CIÊNCIA E TECNOLOGIA  
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO  
CURSO DE BACHAREL EM COMPUTAÇÃO**

**ADRIANO ARAÚJO FELISBERTO**

**ANÁLISE DE DADOS DE MÚSICAS UTILIZANDO A MÍDIA SOCIAL *SPOTIFY***

**CAMPINA GRANDE-PB  
2019**

ADRIANO ARAÚJO FELISBERTO

**ANÁLISE DE DADOS DE MÚSICAS UTILIZANDO A MÍDIA SOCIAL SPOTIFY**

Trabalho de Conclusão de Curso apresentado ao Centro de Ciência e Tecnologia (CCT), Departamento do Curso Ciência da Computação da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Computação.

**Área de concentração:** Computação.

**Orientador:** Prof. Dr. Vladimir Costa de Alencar.

**CAMPINA GRANDE-PB  
2019**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

F315a Felisberto, Adriano Araújo.  
Análise de dados de músicas utilizando a mídia social Spotify [manuscrito] / Adriano Araujo Felisberto. - 2019.  
79 p. : il. colorido.  
Digitado.  
Trabalho de Conclusão de Curso (Graduação em Computação) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2019.  
"Orientação : Prof. Dr. Vladimir Costa de Alencar, Departamento de Computação - CCT."  
1. Consumo de músicas. 2. Streaming. 3. Spotify. I. Título  
21. ed. CDD 600

ADRIANO ARAÚJO FELISBERTO

ANÁLISE DE DADOS DE MÚSICAS UTILIZANDO A MÍDIA SOCIAL SPOTIFY

Trabalho de Conclusão de Curso apresentado ao Centro de Ciência e Tecnologia (CCT), Departamento do Curso Ciência da Computação da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Computação.

Aprovado em: 04 de dezembro de 2019

**BANCA EXAMINADORA**



---

Prof. Dr. Vladimir Costa de Alencar (DC - UEPB)  
Orientador(a)



---

Prof. MSc. Danilo Abreu Santos (DC - UEPB)  
Examinador(a)



---

Profa. MSc. Adília Pires Uchôa Queiroz (IESP)  
Examinador(a)

Aos meus familiares, pela dedicação,  
companheirismo e amizade, DEDICO.

## **AGRADECIMENTOS**

Ao professor Vladimir Alencar pelas leituras sugeridas ao longo dessa orientação e pela dedicação.

A minha mãe, Maria Nazareth e ao meu pai José Isidoro, pelo apoio e ensinamentos ao longo da vida.

À minha noiva, Flávia Silva, por toda compreensão, apoio e incentivo para escolha do tema e durante o desenvolvimento do trabalho.

As minhas tias, em especial à Maria Selma e Maria Salomé, por todo apoio durante o desenvolvimento do trabalho.

Aos colegas de classe pelos momentos de amizade e apoio.

*“A música é a alma da geometria”.*

Paul Claudel

## RESUMO

Neste trabalho, objetivou-se apresentar o seguinte tema: análise de dados de músicas utilizando a mídia social *Spotify*. Nesse sentido, buscou-se, a princípio, mostrar que com a popularização da internet e a evolução tecnológica, surgiram novos meios para o consumo de músicas, que outrora só era possível através de instrumentos musicais, rádios, discos de Vinis, ou LPs, fita cassete e CDs. Assim, surgiram no mercado digital os *streamings* de música, como o *Spotify*, objeto de estudo deste trabalho, o qual possibilita ao usuário ouvir, selecionar as músicas de acordo com o gênero, estilo e época de vários artistas do mundo, podendo ouvi-las em qualquer momento e lugar, bastando ter internet. Nesse sentido, o objetivo geral que norteia este trabalho, é analisar os dados provenientes do *Spotify*. Esses dados são de músicas contidas em quatro países: Brasil, França, Japão e Estados Unidos. A análise consistiu em identificar alguns padrões de estilos tocados nesses países. Para tanto, buscou-se delimitar a metodologia que mais se adequa ao método utilizado, o experimental, sendo assim, a pesquisa classificou-se, em básica, explicativa, bibliográfica, experimental e quantitativa. Como coleta de dados, buscou-se realizar um levantamento bibliográfico sobre a temática nos meios físicos, livros, e na Internet através de artigos sobre o tema utilizando, por exemplo, o Google Acadêmico. Para a fase experimental procurou-se coletar os dados utilizando a biblioteca do *Spotipy*, desenvolvida com a Linguagem de Programação *Python*. As buscas foram feitas por playlists em cada um dos países deste estudo: Brasil, França, Japão e Estados Unidos. Como resultado, obteve-se a quantidade de músicas por país, o estilo, a categoria e os artistas mais ouvidos através de gráficos e tabelas. Os Estados Unidos (US) contêm a maior concentração de músicas, seguido do Brasil. Toda análise foi realizada com base na coleta de dados realizada no período de março de 2019 a junho de 2019. Usando como termo de busca a frase “top 50”, o qual retornou *playlists* com as músicas mais tocadas em cada país. Conclui-se que com a análise foi possível identificar nas *playlists* do *Spotify* diversas músicas e estilos diversificados em cada país estudado, mostrando assim, que o *streaming* se tornou uma importante ferramenta para o consumo de música em qualquer lugar e momento.

**Palavras-Chave:** Consumo de músicas. *Streaming*. *Spotify*. Análise dos Dados.



## ABSTRACT

This paper aims present the theme: data analysis of music using Spotify social media. In this sense, it was sought to find out that with the popularization of the internet and technological evolution, new means to listen to the music emerged, which was only possible through musical instruments, radios, vinyl records (LPs), cassette and CDs. Thus, music streaming, such as Spotify, the object of study of this paper, emerged in the digital Market. Which allows the user to listen to select songs according to the genre, style and era of various artists in the world. Being able to listen to them at any time and place, just having internet. The general objective that guides this paper is analyze data from Spotify, these are from songs contained in four countries: Brazil, France, Japan and the United States. The analysis will find some style patterns played in these countries. To this end, I seek to delimit the methodology that best fits the method used, the experimental. Thus, the research that is presented, is classified as basic, explanatory, bibliographical, experimental and qualitative. As data collection, we sought to conduct a literature review on the subject in physical media, books, and the Internet. Through articles on the subject using, for example, Google Scholar, for the experimental phase sought to collect data using Spotipy's library, developed with Python, it is a programming language. Searches were performed by playlists, in each of the countries in this study. In the analysis and discussion, it shows the number of songs, category and the most listened artists through graphics and table, by countries. The United States, contains the largest concentration of songs, followed by Brazil. All analysis was performed based on data collection from March 2019 to June 2019. Using as a search term the phrase "top 50", which returned playlists of the most played songs in each country. It is concluded that with the analysis it was possible to identify in Spotify playlists various songs and diverse styles in each country studied. Thus, showing that streaming has become an important tool for the consumption of music anywhere and anytime.

**Keywords:** Music consumption. Streaming. Spotify. Data Analysis

## LISTA DE ILUSTRAÇÕES

Figura 1 – Aspecto de um Sistema de Banco de Dados .....	19
Figura 2 – Comparação entre MongoDB e Banco de Dados Relacional .....	21
Figura 3 – Etapa da Mineração de Texto.....	24
Figura 4 – Agrupamento utilizando o K-means.....	27
Figura 5 – Lema do Spotify.....	28
Figura 6 – Características da conta Premium do Spotify.....	29
Figura 7 – Player de música do Spotify .....	30
Figura 8 – Exemplo de arquivo de música no formato JSON.....	34
Figura 9 – Atributos encontrados nas informações da música .....	34
Figura 10 – Exemplo de arquivo de playlist salvo no MongoDB em formato JSON.....	36

## LISTA DE QUADROS

Quadro 1 – Etapas da Mineração de Dados .....	21
Quadro 2 – Algumas definições de Mineração de Texto .....	23
Quadro 3 – Etapas de Mineração de Textos .....	24
Quadro 4 – Natureza dos dados .....	25

## LISTA DE TABELAS

Tabela 1 – Filtragem das informações.....	35
Tabela 2 – Nível de Dançabilidade.....	43
Tabela 3 – Grupos de Andamentos.....	44
Tabela 4 – Nível de Dançabilidade.....	50
Tabela 5 – Características do nível de Dançabilidade.....	57

## LISTA DE GRÁFICOS

Gráfico 1 – Músicas por país .....	37
Gráfico 2 – Média de dançabilidade por país.....	38
Gráfico 3 – Correlação quanto à dançabilidade.....	39
Gráfico 4 – Outra forma de Correlação.....	40
Gráfico 5 – tonalidade mais presente.....	41
Gráfico 6 – Andamentos mais frequentes encontrados na pesquisa.....	42
Gráfico 7 – Preferência de andamentos.....	42
Gráfico 8 – Histograma da Dançabilidade.....	43
Gráfico 9 – Distribuição da dançabilidade – Andamento Lento.....	44
Gráfico 10 – Distribuição da dançabilidade – Andamento Moderado.....	44
Gráfico 11 – Distribuição da dançabilidade – Andamento Rápido.....	45
Gráfico 12 (a) – Artistas mais frequentes – Andamento Lento.....	45
Gráfico 12 (b) – Categorias mais frequentes – andamento Lento.....	46
Gráfico 13 (a) – Artistas mais frequentes – Andamento Moderado.....	46
Gráfico 13 (b) – Categorias mais frequentes – Andamento Moderado.....	47
Gráfico 14 (a) – Artistas mais frequentes – Andamento Rápido.....	47
Gráfico 14 (b) – Categorias mais frequentes – Andamento Rápido.....	48
Gráfico 15 – Artistas mais frequentes.....	49
Gráfico 16 – Artista Johnny Cash – Nível de dançabilidade.....	50
Gráfico 17 – Andamentos – Johnny Cash.....	51
Gráfico 18 – Estilo Musical – Johnny Cash.....	51
Gráfico 19 – Países onde são ouvidas as músicas de Johnny Cash.....	52
Gráfico 20 – Características de Tonalidade – Johnny Cash.....	53
Gráfico 21 – Ed Sheeran – Nível de Dançabilidade.....	54
Gráfico 22 – Estilo musical do artista Ed Sheeran.....	54
Gráfico 23 – Andamentos nas músicas de Ed Sheeran.....	55
Gráfico 24 – Países onde Ed Sheeran é mais escutado.....	56
Gráfico 25 – Tonalidade – Ed Sheeran.....	56
Gráfico 26 – Johann Sebastian Bach – Nível de Dançabilidade.....	57
Gráfico 27 – Estilo musical – Johann Sebastian Bach.....	58
Gráfico 28 – Andamentos – Johann Sebastian Bach.....	58
Gráfico 29 – Países onde mais se escuta Johann Sebastian Bach.....	59

Gráfico 30 – Tonalidade – Johann Sebastian Bach.....	59
Gráfico 31 – Nível de Dançabilidade por país – Andamento Lento.....	60
Gráfico 32 – Níveis de Dançabilidade por país – Andamento Moderado.....	61
Gráfico 33 – Níveis de Dançabilidade – Andamento Rápido.....	61
Gráfico 34 – Artistas mais comuns no Brasil.....	62
Gráfico 35 – Categorias musicais mais comuns no Brasil.....	63
Gráfico 36 – Artistas mais ouvidos na França.....	63
Gráfico 37 – Estilo Musical – França.....	64
Gráfico 38 – Artistas mais ouvidos no Japão.....	65
Gráfico 39 – Estilo musical – Japão.....	65
Gráfico 40 – Artistas ouvidos nos Estados Unidos.....	66
Gráfico 41 – Estilos musicais mais frequentes nos EUA.....	66
Gráfico 42 – O Método Elbow.....	67
Gráfico 43 – Clusterização.....	68
Gráfico 44 – Artistas Cluster Amarelo.....	69
Gráfico 45 – Artistas Cluster Verde.....	70
Gráfico 46 – Artistas Cluster Preto.....	70
Gráfico 47 – Artistas Cluster Azul.....	71

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>16</b>
2.1	Contextualização Musical	16
2.1.1	Andamentos	18
2.2	Conceitos Computacionais	18
2.2.1	Banco de Dados	18
2.2.2	NoSQL	19
2.2.3	MongoDB	20
2.2.4	Mineração de Texto	21
2.2.5	Análise dos Dados	25
2.2.6	Clusterização	26
2.3	Spotify: Música para todos	28
2.4	Estatística Descritiva	31
<b>3</b>	<b>METODOLOGIA</b>	<b>32</b>
3.1	Coleta de Dados	33
3.2	Pesquisa dos dados	33
<b>4</b>	<b>RESULTADOS E DISCUSSÕES</b>	<b>37</b>
4.1	Análise de dados de músicas obtidos da mídia social Spotify	37
4.1.1	Análise Global	37
4.1.2	Análise por Artistas	48
4.1.2.1	Johnny Cash	49
4.1.2.2	Ed Sheeran	53
4.1.2.3	Johann Sebastian Bach	57
4.1.3	Análise de Dados das músicas por País	60
4.1.3.1	Brasil	62
4.1.3.2	França	63
4.1.3.3	Japão	64
4.1.3.4	Estados Unidos	65
4.2	Clusterização: quantidade de cluster através do método <i>Elbow</i>	67
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>73</b>
	REFERÊNCIAS	76

## 1 INTRODUÇÃO

A popularização da internet tornou possível o surgimento de novos meios de consumo. E de acordo com uma pesquisa realizada pelo Instituto Brasileiro de Geografia e Estatística - IBGE, de 2017, dois terços da população do Brasil (69,8%) possuem conexão com a internet.

Com esse crescimento, observa-se que consumir música não é mais como outrora, uma vez que o rádio, em princípio, era o meio mais prático para ouvi-la, mas com o avanço tecnológico e informacional, desde a popularização da internet até os dias atuais, essa ferramenta tornou-se, de certa forma, obsoleta. Visto que novos instrumentos digitais surgiram para dar mais praticidade e poder às pessoas para eleger as músicas que desejam ouvir, diferentemente, do rádio que tem uma programação preestabelecida (COELHO, 2013).

Com o passar do tempo, o modo de consumo de música vai renovando-se, surgindo novos meios de ouvir música, com isso, o usuário pode ter acesso ao que quiser em questão de instantes. Canções que não tocam nas rádios passam a ter vez e até mesmo a ganhar com divulgação instantaneamente (SANTOS; RAMOS, 2016).

Nessa perspectiva, grandes potências no mercado musical surgiram, como por exemplo, o *Spotify*, *Deezer*, *YT Music*, *Google Music*. Todas elas utilizando o *streaming*, como transmissão de dados, esse é uma tecnologia de transferência de dados em tempo real, havendo, então, a necessidade de uma conexão rápida evitando, com isso, travamentos (BRAGA et al, 2016).

Esses aspectos fazem com que as pessoas, atualmente, tenham a possibilidade de escolher as músicas que querem ouvir, quando e onde lhes convêm, pois por meio dos *streamings* supracitados os usuários podem escutar as músicas gravadas e disponibilizadas pelos artista e bandas (MONDELLI et al, 2018). Logo, não estão mais presas a uma programação previamente definida, tornando desse modo, possível a análise e coleta de informações disponibilizadas pelos serviços destas empresas. Além disso, “a coleta e o armazenamento desse tipo de informação é um aspecto importante que surge com a popularização dessas plataformas e que abre espaço para um grande conjunto de possíveis estudos” (MONDELLI et al, 2018).

Desse modo, este trabalho tem como o objetivo realizar uma análise de dados de músicas utilizando a mídia social *Spotify*, esses são de músicas contidas em quatro países: **Estados Unidos, Brasil, França e Japão**. A análise consistirá em identificar



alguns **padrões de estilos** tocados nesses países, nesse sentido, buscou-se na primeira parte do exame, analisar, de maneira global, esses padrões. Demonstrando, com isso, os cantores internacionais mais populares em cada país, podendo então, identificarmos os artistas com maior popularidade e os estilos mais comuns, no âmbito internacional, naquele país. Como também, identificar os artistas em comum dentre os países analisados.

Já a segunda parte da análise consistirá em identificar padrões nas músicas coletadas, empregando para isso, as músicas coletadas em cada país, individualmente. Desse modo, para identificarmos os padrões foi utilizado o critério de **dançabilidade**, mostrando efetivamente, os cantores com nível mais alto de dançabilidade. Identificando, portanto, características com base este atributo. Com isso, o trabalho demonstrará o poder da coleta e análise de dados para possíveis direcionamentos de vendas musicais, anúncios, perfis de consumidores, a mudança na forma de vender música.

Nessa perspectiva, para alcançar esses objetivos, buscou-se delimitar os procedimentos metodológicos para a elaboração da pesquisa. Essa pesquisa configura-se como um conjunto de ações, propostas para encontrar a solução para um problema, as quais têm por base procedimentos racionais e sistemáticos (PRODANOV; FREITAS, 2013). Com isso, a pesquisa, que por ora se apresenta, classifica-se, como básica, explicativa, bibliográfica, experimental e quantitativa.

Enfim, para uma melhor compreensão do tema, o trabalho estrutura-se da seguinte forma: no referencial teórico, o intuito é apresentar, sucintamente, uma contextualização acerca da forma de como as pessoas ouviam música antes dos avanços tecnológicos e dos *streamings*, também, busca-se abordar alguns conceitos e mostrar como a **mídia social** Spotify surgiu. Em seguida, apresenta-se, com mais detalhes, a metodologia aplicada na realização desta pesquisa, por fim, na discussão e resultados, a finalidade é analisar os dados obtidos sobre as músicas mais tocadas no *Spotify* nas regiões selecionadas.

## 2 REFERENCIAL TEÓRICO

Nesta seção, a finalidade é realizar uma breve contextualização sobre a evolução da forma de ouvir música e conceitos básicos sobre andamentos, também, é mostrado como surgiu o *streaming*. Em seguida, apresentam-se alguns conceitos computacionais, bem como, abordam-se algumas características do *Spotify*.

### 2.1 Contextualização Musical

Com o avanço das novas tecnologias e da Internet, ao longo do tempo, a maneira de como as pessoas ouvem músicas e as adquirem mudou de modo considerável. Posto que, se outrora os indivíduos ouviam músicas por meio de rádio, disco de vinil ou por *Compact Disc* (CD), com o surgimento da Internet de banda larga, o consumo de músicas foi facilitado ainda mais. Em primeiro lugar, com o trânsito de arquivos digitais, houve a popularização do mp3 player; depois, com o surgimento dos streamings de músicas, esses instrumentos tornaram-se obsoletos, de acordo com Ribeiro (2019).

Diante desses aspectos, observa-se que a música evoluiu partindo do princípio de técnicas de produção e reprodução, pois antes era produzida por meio de instrumentos musicais, como tambores, por exemplo. Depois, passaram a ser gravadas, no século XX, em virtude dos avanços na eletricidade, e com o descobrimento das ondas hertzianas, século XIX, surgiu o rádio, com isso, foi possível ouvir músicas em casa (SANTOS; RAMOS, 2016).

Em 1940, houve um avanço na maneira de consumir músicas, surgiu o disco de Vinil, os chamados *Long Play* (LP), que as reproduziam através de um material plástico denominado de vinil (feito de PVC), o qual registrava informações de áudio, sendo reproduzidas por meio de toca-discos (TECMUNDO, 2012)<sup>1</sup>. Foi a partir dessa descoberta que a indústria fonográfica surgiu e passou a gravar, produzir, distribuir e divulgar tanto a música quanto os artistas ligados a ela (SANTOS; RAMOS; 2016).

Além disso, com o surgimento das fitas cassetes (K7), em 1963, foi possível a reprodução de música portátil, o que acabou virando uma alternativa para o disco de vinil, permitindo assim, a gravação das músicas do LP para a fita. Essa permitia 30 minutos, de cada lado, de músicas, no entanto, a qualidade do som não era muito

---

<sup>1</sup>Disponível em: <https://www.tecmundo.com.br/infografico/30658-a-evolucao-do-armazenamento-de-musicas-infografico-.htm> acesso em: 24 nov. 2019.

boa. Permitiu também, que as músicas tocadas em rádios pudessem ser gravadas e ouvidas em um aparelho, o *walkman* (COELHO, 2013).

Outra evolução, no modo de ouvir e da distribuição da música, foi com o surgimento do *Compact Disc* (CD), em 1979, desenvolvido, a princípio, para armazenar e tocar música, de forma leve e qualidade superior, com crescimento na durabilidade dos arquivos (SANTOS; RAMOS, 2016). Além disso, o CD tornou mais fácil a forma de ouvir música, pois permitia o armazenamento de 700 MB de dados, cuja qualidade de áudio era mais superior do que a do K7, possibilitando também, a mudança de faixa com rapidez, além de possuir maior vida útil, diferentemente, das fitas magnéticas que eram, com facilidade, corrompidas pelo calor (COELHO, 2013).

Desse modo, com os avanços de aparelhos tecnológicos e da internet de banda larga, surge então o *streaming*, esse constitui-se em uma tecnologia de transferência de dados em tempo real, havendo, então, a necessidade de uma conexão rápida evitando, com isso, travamentos. Não sendo necessário fazer o armazenamento de arquivos no disco rígido (HD) do aparelho, ou seja, realizar o *download* de músicas, tornando assim, mais rápida a acessibilidade a conteúdo *online*, conforme explica Braga et al (2016). Além disso:

Seu modelo de negócio é baseado em uma experiência de consumo de conteúdos digitais que substitui a lógica da compra de um disco pelo acesso a fonogramas hospedados nas redes digitais, permitindo que seu desfrute possa ser realizado sem que se precise baixar, arquivar e organizar esse conteúdo em dispositivos individuais (KISCHINHEVSKY et al, 2019).

Esses aspectos apontados pelos autores são vistos, por exemplo, no *Spotify*, objeto de nosso estudo. Ainda para Kischinhevsky et al (2019), ao utilizar essa tecnologia pela primeira vez, o usuário se depara com uma *interface* intuitiva que apresenta os conteúdos de modo análogo as formas das indústrias fonográfica e radiofônica. Ou seja, o usuário pode escolher para ouvir um fonograma na forma de um álbum, contendo dez ou mais faixas, ou, simplesmente, uma individual (*single*), além disso, pode escutar estações de rádios, essas são organizadas por gêneros musicais, por décadas, tarefas cotidianas entre outras (KISCHINHEVSKY et al, 2019).

Entende-se, desse modo, que o usuário pode escolher e escutar uma diversidade de músicas por meio de aparatos tecnológicos como o celular, computador, notebook. Além disso, poderá também selecionar as músicas por época, estilo, cantores entre outros aspectos. Nesse sentido, observa-se que há um número menor de ouvintes do rádio, pois “é perceptível à busca das pessoas por mobilidade,

conveniência e praticidade. Novas relações sociais são criadas no surgimento de cada produto musical no mercado” (BRAGA et al, 2016).

### **2.1.1 Andamentos**

Andamentos são os movimentos referentes ao movimento rápido ou lento. Os andamentos são divididos em três grupos: lento, moderado e rápido. O primeiro, contém o *Largo*, *Larghetto*, *Lento* e *Adágio*. O moderado contém: *Andante*, *Andantino*, *Moderato* e *Allegretto*. O rápido, *Allegro*, *Vivace*, *Vivo*, *Presto* e *Prestíssimo* (PRIOLLI, 2006).

Portanto, busca-se no seguinte ponto apresentar alguns conceitos computacionais sobre bancos de dados e seus modelos, NoSQL, MongoDB, entre outros aspectos, também, aborda-se acerca do streaming Spotify, apontando as características e sua origem.

## **2.2 Conceitos Computacionais**

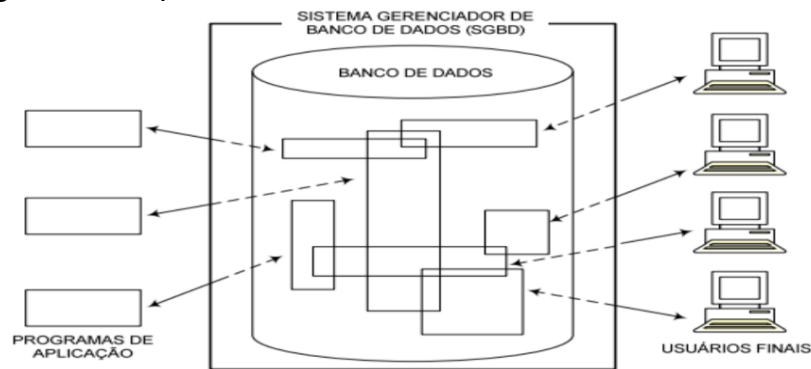
### **2.2.1 Banco de Dados**

De acordo com Esmalri e Navathe (2011), um banco de dados é uma coleção de dados relacionados, ou seja, um relacionamento de informações, em que há um sentido para o usuário. Onde uma tabela representa um objeto do mundo real, que se relaciona com outra tabela. Ainda, um banco de dados representa algum aspecto do mundo real. É uma coleção logicamente coerente de dados. Existe, portanto, uma relação dos dados em que há sentido para um determinado contexto.

Há dois tipos principais de banco de dados, a saber: o primeiro faz referência ao relacional, que é aquele que consiste no relacionamento da informação de fato, muito utilizado no mundo empresarial, e o não-relacional, voltado à grande massas de dados onde existe a necessidade de uma guardar a informação, muitas vezes, sem tratamento. Estes bancos são, geralmente, denominados de NoSQL (SOLAGNA; LAZZARETTI, 2016).

Conforme Esmalri e Navathe (2011), os elementos de um banco de dados são: o próprio banco de dados e o *Software* Gerenciador do Banco de Dados, os que acessam seus componentes, que são os usuários, e suas aplicações (figura 1).

**Figura 1** – Aspecto de um Sistema de Banco de Dados



**Fonte:** DATE, 2004.

De acordo com Solagna e Lazzaretti (2016) citando Date (2004), existe entre o banco de dados armazenado e os usuários um conjunto de programas denominado *Software* Gerenciador de Banco de dados (SGBD), desse modo, entende-se que a gestão dos bancos de dados se dá graças aos SGBDs. Sendo esse um gerenciador responsável pelo gerenciamento do banco de Dados, visto que esse programa dar suporte aos métodos de armazenamento e recuperação das informações adequadamente (SOLAGNA; LAZZARETTI, 2016).

### 2.2.2 NoSQL

O termo *NoSQL* (*Not only SQL*, em português: Não apenas SQL), segundo Suissa (2010, apud SOLAGNA; LAZZARETTI, 2016), foi empregado, primeiramente, por Carlo Strozzi, em 1998, como nome para o seu SGBD (Sistema Gerenciador de Banco de Dados), fundamentado no Modelo Relacional. Eles surgiram como solução para problemas de escalabilidade, disponibilidade e velocidade, encontrados no Modelo Relacional. Pois, com o alto crescimento de dados, através, principalmente, das redes sociais, foi necessário um banco menos burocrático para interação em tempo real do usuário. *NoSQL* suporta, também, a linguagem SQL, porém, rompe o paradigma dos bancos de dados baseados no modelo relacional.

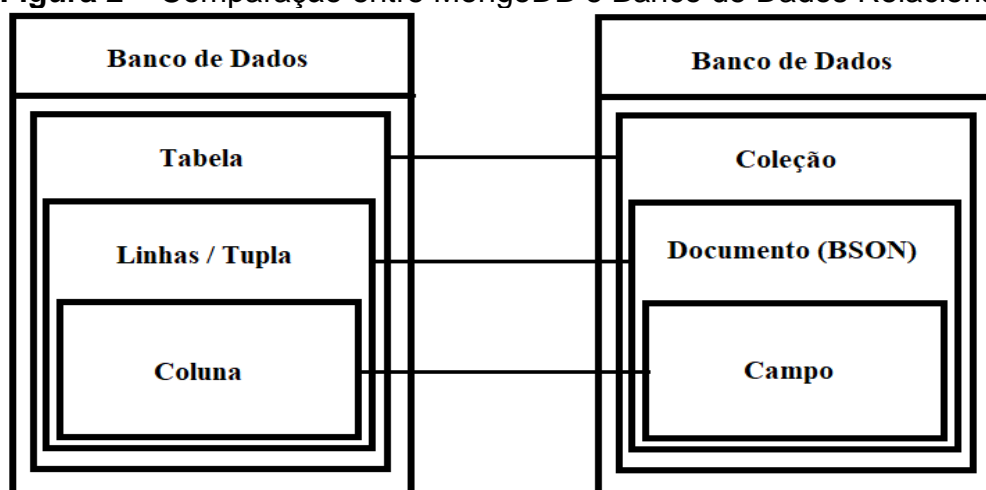
Além disso, o *NoSQL* tem como principais características a maior escalabilidade, volume e processamento. E de acordo com Loscio et al (2011), existe quatro tipos de banco de dados *NoSQL*, são eles:

- Chave-valor: onde o armazenamento é realizado através de um primeiro parâmetro, chamado chave, e o segundo, um valor. Este modelo facilita o acesso à informação, pois cada chave retorna um valor único.
- Orientado a colunas: este tipo de armazenamento é realizado através de valores representativos das colunas. Ou seja, uma coluna é predefinida com valores conhecidos: data\_nascimento, nome, idade. A partir disso, os valores são acessados de forma atômica, mesmo havendo interesse em apenas uma coluna, quando o valor é acessado, retorna a coluna inteira.
- Orientado a Documentos: este modelo armazena os valores como documentos, que possui um identificador único e outros objetos de valores como listas aninhadas, textos. É semelhante com o modelo chave valor, no entanto, esse modelo associa uma única chave a um único valor.
- Orientado a Grafos: este modelo tem três componentes básicos, os nós, os relacionamentos e as propriedades. Podemos identificar este modelo como um multigrafo rotulado e direcionado, onde cada nó pode ser conectado por mais de uma aresta (LOSCIO et al, 2011).

### **2.2.3 MongoDB**

O MongoDB é um banco de dados orientado à documentos, classificado como NoSQL. Pode incluir uma variedade de documentos que são salvos em coleções, gravados em formato BSON, que é uma “extensão do modelo de dados JSON”, em que os valores são mapeados em chave/valor (SOLAGNA; LAZZARETI, 2016). Este modelo é bastante utilizado, atualmente, por permitir uma transferência de informação rápida, pois se trata, basicamente, de um arquivo de texto.

Solagna e Lazzaretti (2016) afirmam que o MongoDB é considerado um dos projetos mais notáveis NoSQL, devido sua linguagem de consulta de alta performance, baseada em documentos e facilidade de migração de um banco de dados relacional. Desse modo, torna possível inserir uma grande quantidade de dados, de diferentes formas e volumes, que é algo primordial no *Big Data*. Também, por estes motivos, o MongoDB se torna escalável, pois não há a necessidade de definir um modelo de relação, visto que a inserção é totalmente feita baseada em documentos, que não têm seus valores fixados. Na figura 1, segue uma visão do mongoDB, em comparação com um Banco de Dados Relacional.

**Figura 2 – Comparação entre MongoDB e Banco de Dados Relacional**

Fonte: <https://www.treinaweb.com.br/blog/sql-vs-nosql-qual-usar/>, 2019. (Adaptada pelo autor).

## 2.2.4 Mineração de Texto

Para alguns autores, como Tan (1999 apud LAZZARIN, 2017), a Mineração de Textos (MT) configura-se como uma extensão do processo de Mineração de Dados (*Data Mining*), conhecida também como Descoberta de Conhecimento em Banco de Dados (*Knowledge Discovery in Databases – KDD*).

A Mineração de Dados constitui-se em uma prática de analisar dados que foram coletados, fazendo uso de vários algoritmos, por vezes, de modo automatizado, cuja finalidade, é a geração de novas informações além de encontrar arquétipos. Sua finalidade é, de acordo com Carvalho Filho (2014) citando Feldman (1999), focar na exploração computadorizada em grandes massas de dados para descobrir padrões interessantes entre elas. Além disso, a maioria dos trabalhos de mineração de dados é realizada em cima de base de dados estruturada (CARVALHO FILHO, 2014). Desse modo, Lazzarin (2017) nos mostra as etapas do KDD, veja-se no quadro 1:

**Quadro 1 – Etapas da Mineração de Dados**

ETAPAS	CARACTERÍSTICAS
<b>DADOS</b>	O KDD se baseia no armazenamento dos dados de forma estruturada
<b>SELEÇÃO DE DADOS</b>	Após ter definido o domínio sobre o qual se pretende executar o processo de descoberta, a próxima etapa é selecionar e coletar o conjunto de dados ou variáveis necessárias

<b>PROCESSAMENTO</b>	Esta etapa é também conhecida como pré-processamento, visando eliminar os dados que não se adequam às informações, com base nos algoritmos, ou seja, dados incompletos, tratamento de nulos, sumarização, problemas de definição de tipos, eliminação de tuplas repetidas, etc.
<b>TRANSFORMAÇÃO</b>	Nesta etapa os dados deverão ser armazenados adequadamente para facilitar na utilização das técnicas de mineração de dados
<b>MINERAÇÃO DE DADOS</b>	A atividade de descoberta do conhecimento é onde são processados os algoritmos de aprendizado de máquina e de reconhecimento de padrões. A maioria dos métodos de Data Mining são baseados em conceitos de aprendizagem de máquina, reconhecimento de padrões, matemática, estatística, classificação, clusterização, modelos gráficos, redes neurais, <i>deep learning</i> , inteligência artificial, etc.
<b>INTERPRETAÇÃO/AVALIAÇÃO</b>	Nesta etapa final, os resultados do processo de descoberta do conhecimento podem ser mostrados de diversas formas, porém devem ser apresentadas de forma que o usuário possa entender e interpretar os resultados obtidos.

**Fonte:** Elaborado pelo autor, 2019.

Nessa perspectiva, os procedimentos de KDD são vistos como práticas que tem por finalidade a melhoria dos resultados das explorações realizadas com o uso de instrumentos clássicos de exploração de dados, ademais, o processo de Descoberta de Conhecimento em Banco de Dados se assemelha ao de Descoberta de Conhecimento em Textos, desde a definição do problema, até a extração de conhecimentos (CARVALHO FILHO).

Por conseguinte, a Mineração de Texto (*Text Mining*) possui variadas terminologias, como por exemplo, *Descoberta de Conhecimento em Textos (Knowledge Discovery from Text – KDT)*, também, *Mineração de Dados Textuais (Text Data Mining)*, e *Descoberta de Conhecimento a partir de Bancos de Dados Textuais (Knowledge Discovery from Textual Databases)*.



## Quadro 2 – Algumas definições de Mineração de Texto

AUTORES	DEFINIÇÃO
Barion et al. (2015, <i>apud</i> LAZZARIN, 2017)	MT pode ser vista como uma extensão da área de <i>Data Mining</i> , focada na análise de textos.
Tan (1999, <i>apud</i> LAZZARIN, 2017)	Mineração de textos: também chamada de <i>Text Data Mining, Knowledge Discovery in Texts (KDT)</i> , refere-se a extração de padrões ou conhecimentos interessantes e não triviais de documentos de texto.
Dorre et al. (1999, <i>apud</i> LAZZARIN, 2017)	Mineração de Textos: se aplica a algumas funções analíticas de mineração de dados, mas também se aplica a funções analíticas de linguagem natural (LN) e técnicas de recuperação de informação - <i>Information Retrival (IR)</i> .
Aggarwal et al. (2012, <i>apud</i> LAZZARIN, 2017)	Mineração de Texto: pode ser considerada como recurso ao acesso à informação onde auxilia os usuários a analisar as informações para a tomada de decisão.

Fonte: Elaborado pelo autor, 2019.

Em síntese, pode-se dizer que Mineração de Texto (*Text Mining*) é um processo que utiliza técnicas de análise de dados, tendo como base textos, para a extração de conhecimentos de dados não estruturados. Esses não possuem qualquer caractere de estrutura definitiva, não tendo, muitas vezes, metadados, caso os tenham não ajuda para um procedimento de análise, por essa razão é que a TM faz relação direta com os dados não estruturados, visto que o texto produz informação e conhecimento de forma explícita, basta lermos o texto para entendermos o seu significado (AMARAL, 2016).

De acordo com Lazzarin (2017) citando os autores Feldman et al (1998), Dorre et al (1999); Tkach (1998) e Wai-Chiu e Fu (2000), os instrumentos de MT são utilizados para as mais variadas funções, tais como:

- Extrair informações relevantes de um documento - extrair as características (entidades) de um documento usando LN, IR e algoritmos métricas de associação (FELDMAN et al, 1998) ou correspondência padrão (DORRE et al.1999);
- Encontrar tendências ou as relações entre pessoas / lugares / organizações, etc, através da agregação e comparar as informações extraídas dos documentos (LAZZARIN, 2017);

- Classificar e organizar documentos de acordo com o seu conteúdo (TKACH,1998);
- Recuperação de documentos com base nos vários tipos de informações sobre o conteúdo do documento (LAZZARIN, 2017).
- Agrupar documentos de acordo com seu conteúdo (WAI-CHIU; FU, 2000).

Na Figura 3, observa-se as etapas do processo de Mineração de Texto apontadas por Carvalho Filho (2014):

**Figura 3 – Etapa da Mineração de Texto**



Fonte: CARVALHO FILHO, 2014.

Esse modelo de etapas de MT foi proposto por Aranha (2007) citada por Carvalho Filho (2014), assim, em suma veja-se como essas etapas funcionam:

**Quadro 3 – Etapas de Mineração de Textos**

ETAPAS	ATRIBUTOS
COLETA	Utiliza-se <i>web crawlers</i> , programas que visitam sítios e extraem informações, para extração dos textos que serão utilizados para a extração de conhecimento. (CARVALHO FILHO, 2014).
PRÉ-PROCESSAMENTO	São utilizadas técnicas como o Processamento de Linguagem Natural para estruturar os textos que serão analisados. (CARVALHO FILHO, 2014).
INDEXAÇÃO	É a etapa onde são extraídos conceitos dos documentos através da análise de seu conteúdo e traduzidos em termos da linguagem de indexação. Esta representação identifica o documento e define seus pontos de acesso para consultas (GOMES, 2006 apud CARVALHO FILHO, 2014)

MINERAÇÃO	São aplicados métodos e algoritmos para a identificação de padrões interessantes e extração de conhecimento
ANÁLISE	Os resultados são avaliados e validados (CARVALHO FILHO, 2014).

**Fonte:** Elaborado pelo autor, 2019.

Por fim, vale ressaltar que no pré-processamento, conforme Amaral (2016), são executadas a remoção das *stop words*, essas são palavras que não possui valor semântico ao processo de mineração. As *stop words* podem variar conforme a finalidade da Mineração. Outras operações são: *steming*, que é a extração dos radicais de palavras, de maneira que essas sejam agrupadas juntas; remoção de números, pontuação, acentuação, espaços em branco, caracteres gráficos, *emoticons*, entre outros aspectos.

### 2.2.5 Análise de dados

De acordo com Amaral (2016), a análise de dados se configura como aplicação de algum tipo de mudança nos dados para buscar conhecimento. A finalidade da análise é realizar a descrição e resumir os dados, bem como, identificar as relações e diferenças entre variáveis, fazer a comparação dessas e, por fim, realizar previsões.

Para Morais (2005), o método adequado para interpretação dos dados está sujeito à sua natureza, e distingue duas tipologias básicas: dados qualitativos e dados quantitativos.

#### Quadro 4 – Natureza dos dados

DADOS QUALITATIVOS	DADOS QUANTITATIVOS
<ul style="list-style-type: none"> <li>Os dados qualitativos representam a informação que identifica alguma qualidade, categoria ou característica, não susceptível de medida, mas de classificação, assumindo várias modalidades.</li> <li>Para sumariar dados qualitativos numericamente utilizam-se, entre outras formas, contagens, proporções, percentagens, taxas por 1000, taxas por 1000000, dependendo da escala apropriada (MORAIS, 2005).</li> </ul>	<ul style="list-style-type: none"> <li>Os dados quantitativos representam informação resultante de características susceptíveis de serem medidas, apresentando-se com diferentes intensidades, que podem ser de natureza discreta (descontínua) ou contínua.</li> <li>Podem ser discretos ou contínuos. (MORAIS, 2005).</li> </ul>

**Fonte:** Elaborado pelo autor, 2019.

Amaral (2016) assinala dois tipos de análises: a implícita, em que as informações não são claras no conjunto de dados, é mais complexa. Nesse caso, será preciso o uso de alguma função sofisticada para que produza as informações, assim, a dica é a utilização de tarefa de aprendizado de máquina ou uma lei estatística. Já a análise explícita, como a denominação sugere, as informações e o conhecimento estão dispostos de forma clara nos dados, sendo apenas necessário realizar operações de baixa complexidade para observar o dado e criar a informação. A exemplo, aplicar um filtro, em *drill down* num cubo; ordenar registros; criar colunas calculadas ou instrução SQL.

Segundo Amaral (2016), é imprescindível categorizar as análises de dados, visto que se constitui num procedimento em busca de técnica ou instrumento adequado, conforme a finalidade da análise. Entretanto, vale ressaltar que muitas vezes uma mesma técnica ou ferramenta pode ser utilizada com objetivos diferentes.

A Análise Exploratória de Dados, proposta pelo estatístico John Wilder Tukey (1977), tem como finalidade o conhecimento dos dados antes da tentativa de realizar a análise para, em seguida, usar técnicas explícitas ou implícitas para, só assim, poder extrair conclusões dos dados obtidos. A Análise Exploratória faz uso tanto de técnicas quantitativas quanto das técnicas visuais, exemplos da primeira, são as medidas de dispersão e posição como média, mediana, amplitude e desvio padrão (AMARAL, 2016).

### **2.2.6 Clusterização**

A habilidade de classificar e agrupar é algo inerente ao homem desde cedo, através de classificação de coisas, animais, plantas, etc (EVERITT ET AL, 2011). Por exemplo, segundo Grus (2016), um conjunto de dados demográficos de eleitores registrados forma agrupamentos por mães de praticantes de futebol, aposentados, entre outros grupos de pessoas ou características relevantes. Segundo Everitt et al (2011), a classificação é necessária para o desenvolvimento da linguagem. Nesse sentido, poderá ser dado nome as coisas, classificando-as. Estamos, portanto, falando de clusterização, ou agrupamento.

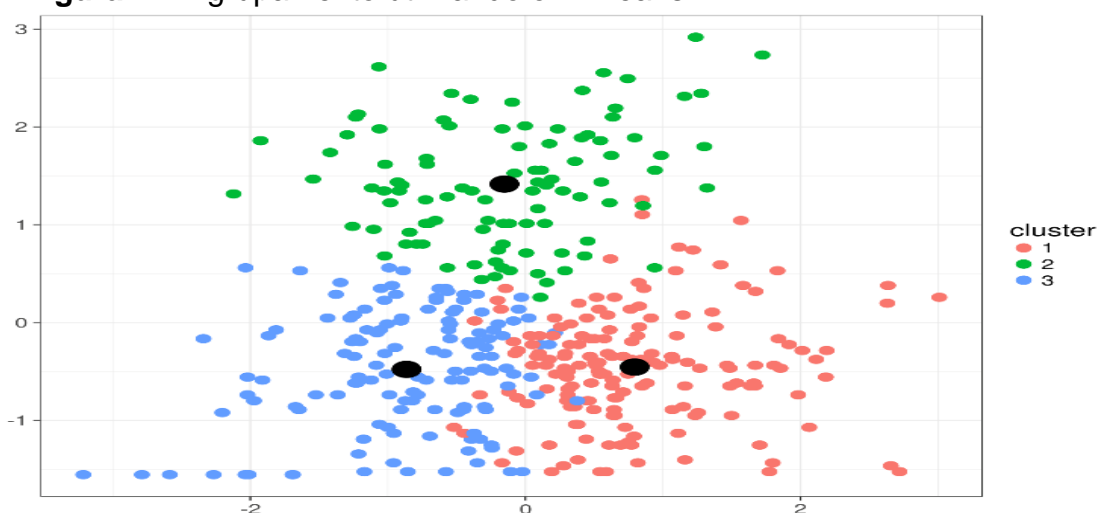
Clusterizar consiste em métodos de criar grupos de objetos no qual possuem muitas similaridades, de acordo com Gan et al (2007). A classificação pode envolver

peças, animais, plantas, elementos químicos, estrelas, entre vários outros (EVERITT ET AL, 2011).

Existem vários tipos de algoritmos de agrupamentos. A exemplo o *K-means*, é baseado em centroides, no qual o número de agrupamentos é definido pelo usuário, (AMARAL, 2016). Além disso, para Benito et al (2017), o *K-means clustering* configura-se como um método que tem por base a distância, a qual consiste em criar clusters para que a variação total intra-cluster se torne mínima.

Na Figura 4, mostra-se um agrupamento, utilizando o K-means.

**Figura 4** – Agrupamento utilizando o K-means



Fonte: <https://rpubs.com/cyobero/k-means> Acesso em 16 nov. 2019.

Pode-se observar que o agrupamento da Figura 7 foi feito em 3 grupos, como mostrado na legenda a direita da figura. Os centroides são os pontos pretos, no centro dos grupos, divididos pelas cores: vermelha, cluster 1; verde, cluster 2; azul, cluster 3.

Para Benito et al (2017), o método de cluster lida com a localização de afinidades e estrutura numa coletânea de pontos de dados, ainda para os autores “no agrupamento k-means, cada agrupamento é representado por seu centro (isto é, centroide), que corresponde à média de pontos atribuídos ao agrupamento”<sup>2</sup>. Assim, os centroides são os centros dos clusters, também, referenciados como método de agrupamento de pares não ponderados usando centroides, segundo Gan et al (2011), citando Jain e Dubes (1988).

<sup>2</sup> Tradução nossa. “In k-means clustering, each cluster is represented by its center (i.e, centroid) which corresponds to the mean of points assigned to the cluster” (BENITO et al, 2017).

A princípio, os centroides são atribuídos de forma aleatória. É iniciada iterações para identificar os dados mais próximos a cada ponto, chamados de vizinhos mais próximos, onde os resultados obtidos são armazenados em uma matriz. Conforme a iteração vai acontecendo, os vizinhos são modificados, com isso, é necessário recalcular os centroides. A cada iteração os centroides são modificados (EVERITT et al, 2011). E de acordo com Lazzarin (2017),

Na descoberta de conhecimento em textos, no lugar de atributos de uma tabela de banco de dados relacionais, utilizam-se estatísticas de aparições de termos ou frases nos documentos. Os relacionamentos entre termos lexicamente diferentes, porém semanticamente similares, também podem ser considerados para melhorar os resultados da técnica (LAZZARIN, 2017).

Enfim, buscou-se abordar alguns aspectos da clusterização de forma sucinta, mais adiante, nos resultados, será analisada com mais ênfase para mostrar como foram agrupadas as músicas por nível de dançabilidade e positividade, e a escolha da quantidade de cluster foi realizada por meio do método *Elbow*.

### 2.3 Spotify: Músicas para todos

Com lema “Música para todos”, o Spotify configura-se como um dos maiores serviços de *streaming* para o consumo de músicas, comportando uma variedade de gêneros musicais e de vários países em um único lugar, sendo disponível a todo o público. A plataforma possui aproximadamente 30 milhões de músicas em seu catálogo (SANTOS; RAMOS, 2016).

**Figura 5** – Lema do Spotify



Fonte: Print Screen Lightshot. Site: <https://promo.spotify.com/>, 2019.

A empresa sueca foi desenvolvida, em 2006, por Daniel Ek e Martin Lorentzon, mas foi lançada apenas em 2008 pela *Spotify* AB, que tem sede em Estocolmo, chegando ao Brasil somente em 2014. A plataforma é disponibilizada para *download* para computadores, tablets e celulares, sendo que é necessário ter acesso à internet, uma vez que as músicas podem ser acessadas pelo indivíduo via serviço de *streaming* (BRAGA et al, 2016).

Para acessar o *Spotify*, é preciso baixar o aplicativo e ter uma conta, podendo se inscrever utilizando e-mail e senha ou através da rede social, *Facebook*. São duas possibilidades para adquirir o serviço, uma conta grátis, a qual o usuário pode ouvir músicas com certas restrições, como por exemplo, anúncios durante a reprodução, ou usuário poderá optar pela conta *premium*, a qual obtém benefícios, como ouvir música sem estar conectado à internet. Mas também, oferece três meses grátis para o usuário experimentar a conta premium, sem compromisso, podendo desvincular-se antes de acabar o prazo da gratuidade. O *Spotify*, com o lema “*Premium* do jeito que você quiser”, disponibiliza quatro opções desse tipo de conta e os benefícios de aderir-la, veja-se na figura 6:

**Figura 6** – Características da conta *Premium* do *Spotify*

Individual	Duo	Família	Universitário
3 meses grátis	3 meses grátis	3 meses grátis	3 meses grátis
R\$ 16,90/mês, após o teste	R\$ 21,90/mês, após o teste	R\$ 26,90/mês, após o teste	R\$ 8,50/mês
1 conta	2 contas	6 contas	1 conta
<ul style="list-style-type: none"> <li>✓ Curta música sem anúncios</li> <li>✓ Ouça em qualquer lugar – até offline</li> <li>✓ Toque o que quiser</li> </ul>	<ul style="list-style-type: none"> <li>✓ Curta música sem anúncios</li> <li>✓ Ouça em qualquer lugar – até offline</li> <li>✓ Toque o que quiser</li> <li>✓ Duo Mix: uma playlist diária especial para os dois pombinhos</li> </ul>	<ul style="list-style-type: none"> <li>✓ Curta música sem anúncios</li> <li>✓ Ouça em qualquer lugar – até offline</li> <li>✓ Toque o que quiser</li> <li>✓ Para familiares que moram no mesmo endereço</li> </ul>	<ul style="list-style-type: none"> <li>✓ Curta música sem anúncios</li> <li>✓ Ouça em qualquer lugar – até offline</li> <li>✓ Toque o que quiser</li> <li>✓ 50% de desconto para estudantes universitários aprovados</li> </ul>
INICIAR TESTE GRÁTIS	INICIAR TESTE GRÁTIS	INICIAR TESTE GRÁTIS	COMEÇAR A USAR GRÁTIS

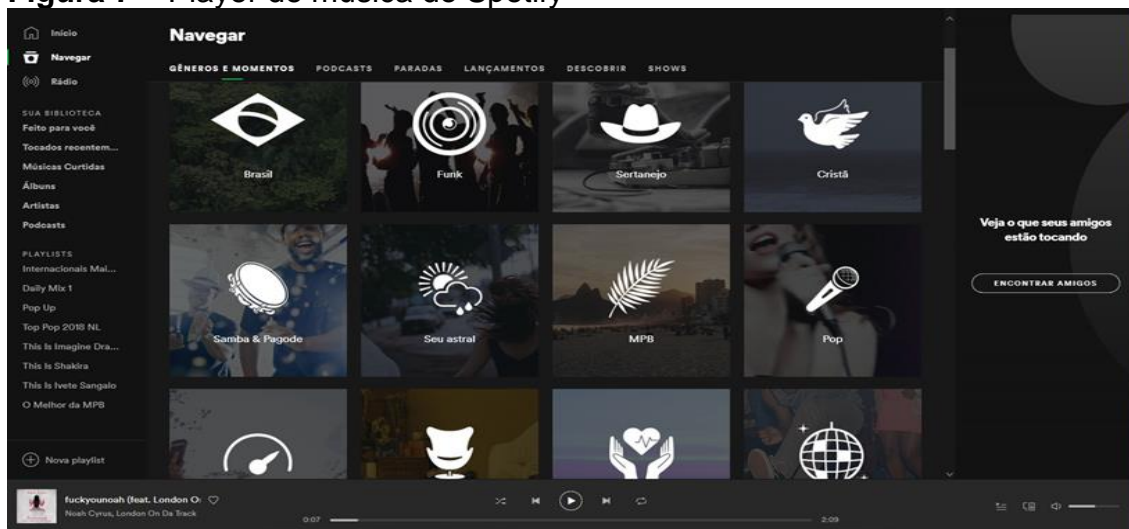
Fonte: Print Screen Lightshot. Site: <https://promo.spotify.com/>, 2019.

O *Spotify* faz uso dos dados em quase todas as suas extensões, pois constitui-se em um alicerce para a tomada de decisões, sempre que for possível, e apenas com o desenvolvimento de técnicas e ferramentas de *Big Data*, o *Spotify* pode realizar tudo o que faz (Figura 10). Além disso, o amplo volume de dados produzido pelos seus usuários abre diversas oportunidades, fazendo com que a empresa possa gerar valor



por meio de sua análise, bem como, utilizá-los para diversas finalidades dentro da própria empresa (RODRIGUES et al, 2016).

**Figura 7** – Player de música do Spotify



**Fonte:** Print Screen Lightshot. Aplicativo Spotify, 2019.

Como pode-se observar na Figura 7, o *Spotify* apresenta diversas funcionalidades como player de música, possibilitando assim, que os usuários possam salvar álbuns, artistas e *playlists* conforme estilo musical, como também, por época, para ouvir quando quiser. De acordo com Braga et al (2016), essas *playlists* podem ser elaboradas por quem usa esse serviço ou pelos curadores, que as sugerem ao usuário, além disso, apresenta um catálogo de músicas variadas, que podem ser buscada tanto pelo nome do artista, pelo estilo quanto pelo álbum. Há, também, a possibilidade de acessar o catálogo por gênero, ouvir *podcasts* e rádio, entre outros aspectos.

Além disso, a possibilidade de escutar rádio dentro do *Spotify* constitui-se em um diferencial, pois as músicas são separadas por gêneros musicais e estão dentro do serviço de modo que o usuário acesse variadas músicas, sem, no entanto, terem sido selecionadas por ele ou não. Nessa perspectiva, com esse grande volume de opções que a plataforma oferece, há um questionamento quanto aos direitos autorais das músicas, mas, com “as publicidades que, teoricamente, mantêm todo o sistema, o Spotify se responsabiliza a pagar pelos conteúdos dos artista”, isso por meio de uma Gestão de Direitos Autorais (GDD), Daniel Ek e Martin Lorentzon garantem que é a própria empresa que gera os direitos sobre as canções, entre outros aspectos (SANTOS; RAMOS, 2016).



Logo, compreende-se que o *Spotify* é uma plataforma que comporta diversas opções de escutar músicas, e apresenta uma interface interativa em que o usuário pode realizar ações, como criar playlists, compartilhar e interagir com outros usuários.

## **2.4 Estatística descritiva**

A Estatística é vista como a ciência dos dados, esses se configuram como instrumentos basilares para a compreensão dos fatos e fenômenos que nos cercam. Para o registro e o estudo da coleta de dados, comumente, é usado a Estatística descritiva que se constitui em um conjunto de técnicas analíticas utilizadas para resumir o conjunto dos dados recolhidos numa dada investigação, que são organizados, geralmente, através de números, tabelas e gráficos (MORAIS, 2005).

De acordo com Ferreira (2005), a estatística descritiva tem como finalidade descrever os dados coletados tanto de amostra quanto de uma população, podendo assim, incluir: verificação da representatividade ou da falta de dados; ordenação dos dados; compilação dos dados em tabela; criação de gráficos com os dados; calcular valores de sumário, tais como médias, medianas e desvio-padrão; e obter relações funcionais entre variáveis.

Portanto, buscou-se nesta seção apresentar alguns conceitos fundamentais sobre bancos de dados, NoSQL, MongoDB, e também mostrar algumas características do streaming de música Spotify, com a finalidade de entender como se processa esses aspectos na fase experimental, em que se busca analisar os dados de músicas obtidos do streaming. Adiante, apresenta-se os procedimentos metodológicos aplicados nesta pesquisa.

### 3 METODOLOGIA

No contexto científico, a pesquisa é classificada conforme critérios, que podem ser quanto à natureza, aos objetivos e aos procedimentos técnicos. De acordo com Wazlawick (2010), “nem sempre um trabalho de pesquisa limita-se a um único tipo. Além disso, alguns tipos de pesquisa podem servir de base para outros.” Nesse sentido, de acordo com essa tipologia, buscou-se delimitar, neste trabalho, o procedimento metodológico mais adequado para o tema abordado.

Nessa perspectiva, quanto à natureza, a pesquisa classifica-se como básica, pois “objetiva gerar conhecimentos novos úteis para o avanço da ciência sem aplicação prática prevista. Envolve verdades e interesses universais (PRODANOV; FREITAS, 2013). Quanto aos objetivos, a pesquisa pode ser exploratória, descritiva ou explicativa, nesse caso, esta pesquisa configura-se como explicativa, visto que nesse tipo procura-se elucidar os porquês dos fatos e suas causas através de registro, da análise, da classificação dos fenômenos observados. Além disso, “é a pesquisa científica por excelência porque, além de analisar os dados observados, ela busca suas causas e explicações, ou seja, os fatores determinantes destes dados” (WAZLAWICK, 2010).

Em relação aos procedimentos técnicos, pode-se classificar esta pesquisa como bibliográfica e, também, experimental. Desse modo, entende-se que “a pesquisa bibliográfica é um passo fundamental e prévio para qualquer trabalho científico, mas ela, em si, não produz qualquer conhecimento novo. Ela apenas supre ao pesquisador informações públicas que ele ainda não possuía” (WAZLAWICK, 2010).

É experimental porque ao determinar um objeto de estudo, pode-se selecionar as variáveis que seriam capazes de influenciá-lo, bem como, é definido as formas de controlar e de observar os efeitos que a variável produz no objeto. Ainda, “a pesquisa experimental é mais frequente nas ciências tecnológicas e nas ciências biológicas. Tem como objetivo demonstrar como e por que determinado fato é produzido” (PRODANOV; FREITAS, 2013). Quanto à abordagem do problema, a pesquisa se configura como quantitativa, pois faz uso da estatística para interpretar os dados coletados.

### 3.1 Coleta dos Dados

Para a fase da pesquisa bibliográfica, buscou-se realizar um levantamento bibliográfico que se aproximasse do tema abordado neste estudo. Sendo assim, através dos termos streamings de música, *Spotify*, Bancos de dados, foi possível encontrar artigos na Internet, usando a ferramenta do Google Scholar e outras, como exemplo, “Um estudo comparativo entre o MongoDB e o PostgreSQL”, de Solagna e Lazzaretti (2016); em meio físico, através de livros como “Sistemas de banco de dados”, das autoras Elmasri e Navathe (2011).

Já para a fase experimental, buscou-se tratar dos dados obtidos da biblioteca do Spotify, por meio das playlists tocadas nos seguintes países: **Estados Unidos, Brasil, França e Japão**. Os resultados da análise exploratória dos dados foram dispostos e organizado em gráficos e tabelas com a finalidade de realizar a interpretação dos dados coletados.

### 3.2 Pesquisa dos dados

A pesquisa dos dados foi realizada utilizando a biblioteca do *Spotipy*, desenvolvida com a Linguagem de Programação *Python*. As buscas foram feitas por *playlists*, em cada um dos países deste estudo: **Brasil, França, Japão e Estados Unidos**. As informações coletadas foram salvas no MongoDB, no formato JSON. Nesta primeira pesquisa foram encontradas 5273 *playlists*. Na Figura 3, é mostrado um exemplo do arquivo, com os atributos de uma *playlist* salvo no MongoDB em formato JSON.

Após a coleta das *playlists* dos países, foi realizada uma coleta referente às músicas contidas nas *playlists*, nas quais foram encontradas 67.439 músicas, contendo músicas repetidas. Após o tratamento das músicas, removendo valores não identificados e músicas repetidas, a quantidade foi reduzida para **40.768 músicas**. A Figura 8 mostra o exemplo de um arquivo de música no formato JSON.

**Figura 8** – Exemplo de arquivo de música no formato JSON

```
{
  'id': ObjectId('5dbd89f7668e5d02299a49b8'),
  'disc_number': 1,
  'duration_ms': 157928,
  'episode': False,
  'explicit': False,
  'external_ids': {
    'isrc': 'BR6RI1900041'
  },
  'external_urls': {
    'spotify': 'https://open.spotify.com/track/29caulxU0jxucdVIyhSJTc'
  },
  'href': 'https://api.spotify.com/v1/tracks/29caulxU0jxucdVIyhSJTc',
  'id': '29caulxU0jxucdVIyhSJTc',
  'is_local': False,
  'name': 'Amor de Que',
  'popularity': 0,
  'preview_url': 'https://p.scdn.co/mp3-preview/9b67922179c9852bb8d45c7b3e810edeab401e7b?cid=868451cf9ab6450b90d202d396a7abcb',
  'track': True,
  'track_number': 2,
  'type': 'track',
  'uri': 'spotify:track:29caulxU0jxucdVIyhSJTc',
  'country': 'BR',
  'category': 'toplists'
}
```

**Fonte:** Elaborada pelo autor, 2019.

Com as informações da música, foi possível coletar informações necessárias para a análise. A Figura 9 mostra os atributos encontrados.

**Figura 9** – Atributos encontrados nas informações da música

```
{'danceability': 0.737,
 'energy': 0.819,
 'key': 11,
 'loudness': -5.056,
 'mode': 0,
 'speechiness': 0.0446,
 'acousticness': 0.0726,
 'instrumentalness': 0,
 'liveness': 0.36,
 'valence': 0.904,
 'tempo': 139.988,
 'type': 'audio_features',
 'id': '29caulxU0jxucdVIyhSJTc',
 'uri': 'spotify:track:29caulxU0jxucdVIyhSJTc',
 'track_href': 'https://api.spotify.com/v1/tracks/29caulxU0jxucdVIyhSJTc',
 'analysis_url': 'https://api.spotify.com/v1/audio-analysis/29caulxU0jxucdVIyhSJTc',
 'duration_ms': 157929,
 'time_signature': 4}
```

**Fonte:** Elaborada pelo autor, 2019.

Estas informações foram filtradas, de acordo com a necessidade da pesquisa, que teve como atributo principal a **dançabilidade** (*danceability*). Os atributos de **energia** (*energy*) e **positividade** (*valence*), foram utilizados pela alta correlação com a **dançabilidade**.

A **dançabilidade** é o quão dançável a música é. É uma combinação de ritmos, andamentos e regularidade. É um atributo que varia de 0 a 1.0. Sendo o nível mais baixo a representação de uma música menos dançante. Já o nível mais alto, o maior nível, representa uma música dançante.

A **energia** representa a intensidade e atividade da música, é uma medida entre 0.0 e 1.0, a qual a menor representa baixa atividade e intensidade na música. O nível

mais alto, representa uma música muito intensa, envolvendo muito atividade, movimento. Tem relação com o nível de **dançabilidade**, conforme a música é dançante, mais movimentos são realizados.

A **positividade** é uma medida entre 0 e 1, a qual representa o quão a música é feliz, alegre. A euforia que a música pode trazer. Possui relação com o nível de **dançabilidade**, ou seja, quanto mais dançante a música é, mais positiva ela pode ser.

O **tempo** é uma medida do tempo, estimado em batidas por minuto. A batida está diretamente correlacionada com o **Ritmo**, ou andamento, que é subdividido em três grupos, para a análise. Os grupos de ritmos são: lentos, moderados e rápidos. Cada grupo possui um andamento relacionado. Para os ritmos lentos, os andamentos são: largo e adagio. Para os ritmos moderados, o andamento é o andante. Para os ritmos rápidos, os andamentos são: allegro, vivace e presto.

O **tom** é a representação da tonalidade da música, utilizando como base a escala musical. O **modo** é a representação da modalidade da música. Maior, para músicas com tonalidade maior. Menor, para as músicas com tonalidade menor.

A **popularidade** é o nível de popularidade de cada música, é medida entre 0 e 100. Onde o maior número significa que a música, é muito popular.

**Tabela 1 – Filtragem das informações**

Artista	Estilo	Pais	Duração (ms)	Nome	Popularidade	Dançabilidade	Energia	Vivacidade	Positividade	Tempo	Ritmo	Tom	Modo
Marília Mendonça	sertanejo	BR	147748	Supera - Ao Vivo	85	0,665	0,743	0,959	0,658	131,573	Allegro	E	Maior
Gusttavo Lima	sertanejo	BR	215766	Carreira Solo - Ao Vivo	0	0,806	0,933	0,871	0,916	126,957	Allegro	D	Maior
Luan Santana	sertanejo	BR	180001	choque térmico - ao vivo	80	0,737	0,758	0,834	0,734	126,978	Allegro	D	Maior
Jorge & Mateus	sertanejo	BR	181930	Cheirosa - Ao Vivo	79	0,528	0,87	0,333	0,714	152,37	Vivace	B	Maior
Anitta	sertanejo	BR	194771	Some que ele vem atrás	68	0,648	0,795	0,38	0,598	180,043	Vivace	C	Menor
Israel & Rodolfo	sertanejo	BR	156838	Também Sei Fazer Falta - Ao Vivo	78	0,715	0,761	0,684	0,783	151,932	Vivace	C	Maior
Zé Neto & Cristiano	sertanejo	BR	159164	Ferida Curada	84	0,49	0,845	0,213	0,548	146,105	Vivace	D	Maior
Maiara & Maraisa	sertanejo	BR	173891	Aí Eu Bebo - Ao Vivo	69	0,827	0,683	0,409	0,85	135,944	Allegro	B	Menor

**Fonte:** Elaborada pelo autor, 2019.

**Figura 10** – Exemplo de arquivo de playlist salvo no MongoDB em formato JSON

```

{
  "_id": ObjectId("5dbd875b668e5d02299a348e"),
  "collaborative": false,
  "external_urls": {
    "spotify": "https://open.spotify.com/playlist/37i9dQZEVXbM0kSwG072hV"
  },
  "href": "https://api.spotify.com/v1/playlists/37i9dQZEVXbM0kSwG072hV",
  "id": "37i9dQZEVXbM0kSwG072hV",
  "images": [
    {
      "height": null,
      "url": "https://charts-images.scdn.co/VIRAL_BR_DEFAULT.jpg",
      "width": null
    }
  ],
  "name": "Brazil Viral 50",
  "owner": {
    "display_name": "spotifycharts",
    "external_urls": {
      "spotify": "https://open.spotify.com/user/spotifycharts"
    },
    "href": "https://api.spotify.com/v1/users/spotifycharts",
    "id": "spotifycharts",
    "type": "user",
    "uri": "spotify:user:spotifycharts"
  },
  "primary_color": null,
  "public": null,
  "snapshot_id": "NjI1OTYyNzY5LDAwMDAwMDAwMDAwMDAwMDAwMDAwMDAwMDAwMDAwMDAwMDY1NmU=",
  "tracks": {
    "href": "https://api.spotify.com/v1/playlists/37i9dQZEVXbM0kSwG072hV/tracks",
    "total": 50
  },
  "type": "playlist",
  "uri": "spotify:playlist:37i9dQZEVXbM0kSwG072hV",
  "country": "BR",
  "category": "toplists"
}

```

**Fonte:** Elaborada pelo autor, 2019.

Portanto, buscou-se com essa metodologia realizar a análise dos dados de músicas da mídia social *Spotify*. Nesse sentido, no próximo capítulo apresenta-se, a priori, como foi realizada a pesquisa dos dados, com mais detalhes, mostrando as músicas mais frequentes nos países selecionados para este estudo. A posteriori, aborda-se a análise propriamente dita, ou melhor, o foco principal da pesquisa, apontando por meio de gráficos os artistas, as categorias de músicas, a dançabilidade entre outros aspectos.

## 4 RESULTADOS E DISCUSSÕES

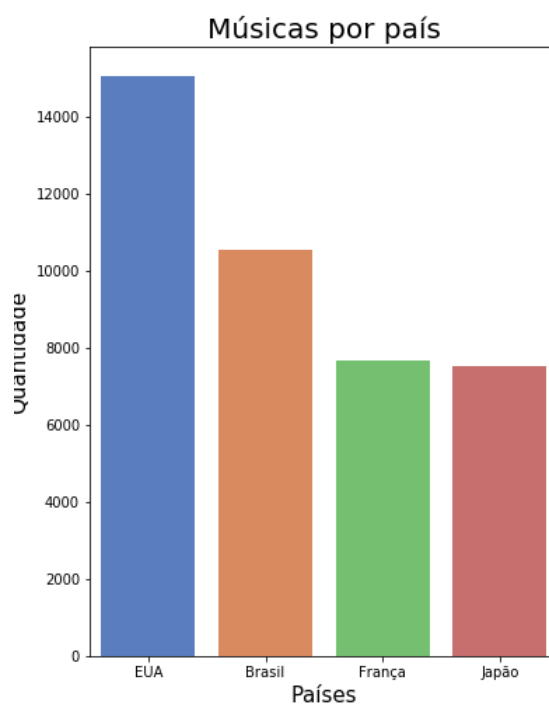
Nesta seção, será apresentada as etapas da pesquisa, contendo informações sobre a análise dos principais dados obtidos, os resultados encontrados e mostrar o que podemos resolver com os dados.

### 4.1 Análise de dados de músicas obtidos da mídia social *Spotify*

#### 4.1.1 Análise Global

No Gráfico 1, é mostrado a quantidade de músicas por país. Os Estados Unidos (US) contêm a maior concentração de músicas, que é 14.000 seguido do Brasil com 10.000 músicas.

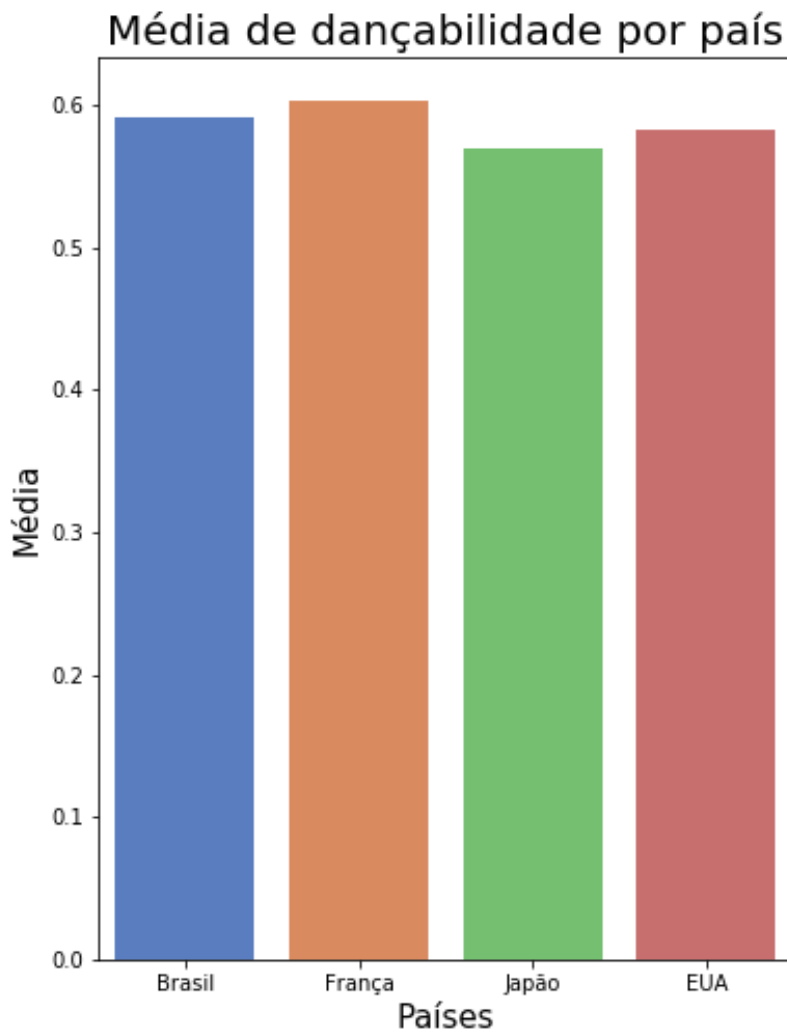
**Gráfico 1** – Músicas por país



**Fonte:** Elaborado pelo autor, 2019.

Apesar de ter uma quantidade de músicas bem mais relevante, a média da dançabilidade, para cada país, é semelhante, como mostrado no Gráfico 2.

**Gráfico 2** –Média de dançabilidade por país.



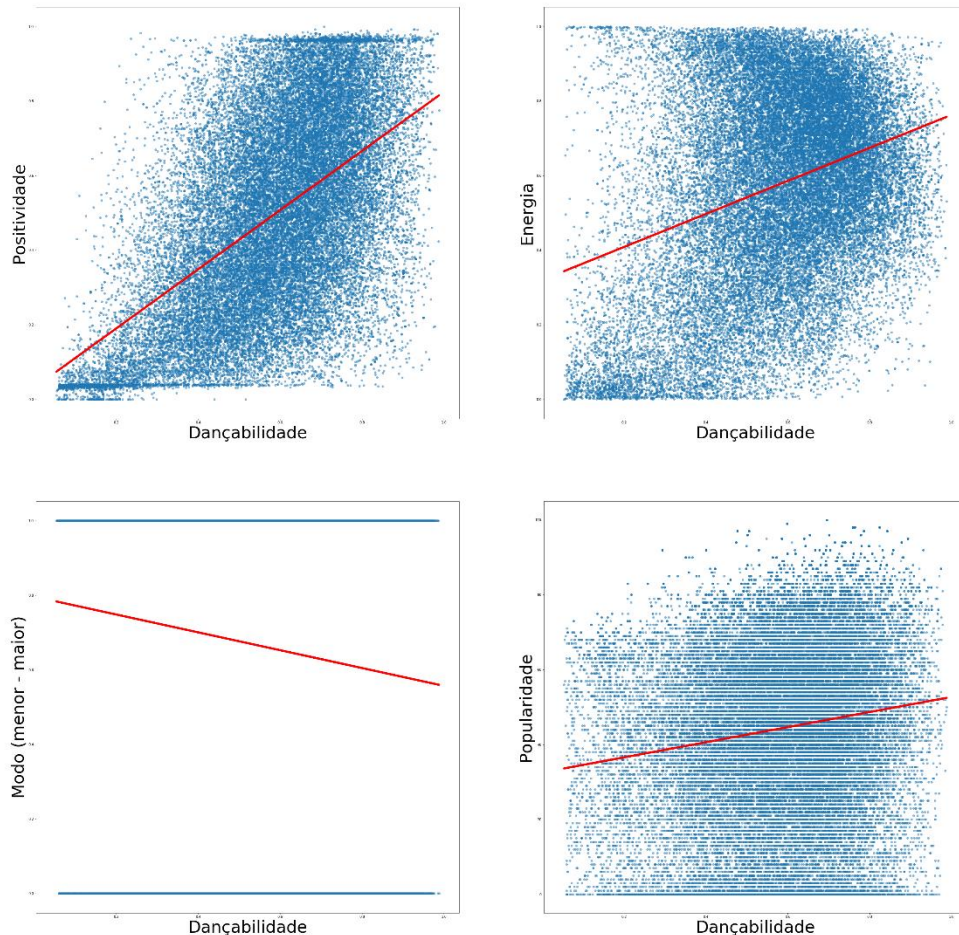
**Fonte:** Elaborado pelo autor, 2019.

No Gráfico 3, é mostrada a correlação entre as variáveis utilizadas, comparando com a dançabilidade, sendo utilizado um gráfico de dispersão dos dados.



### Gráfico 3 – Correlação quanto à dançabilidade

#### Correlação em relação a dançabilidade



**Fonte:** Elaborado pelo autor, 2019.

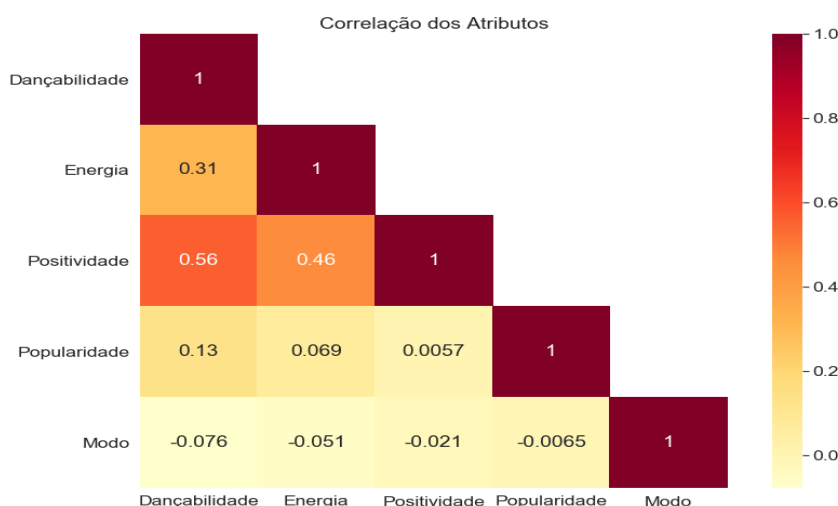
Uma informação importante em relação a este gráfico, é que a linha vermelha demonstra o nível de correlação. Quanto mais inclinado para cima, temos uma correlação maior. Portanto, analisando cada correlação, o nível de dançabilidade correlacionado com o nível de positividade da música, temos uma linha mais inclinada de forma crescente. Dessa forma, podemos concluir que quanto maior o nível de dançabilidade, maior será o nível de positividade da música, assim como a energia.

Já a correlação com o modo da música, se é maior ou menor, há uma correlação no sentido de quando maior a dançabilidade, o modo da música se torna menor. Apesar de, geralmente na música, a tonalidade menor é voltada para músicas

mais tristes, mas a popularidade da música não tem tanta correlação. Portanto, a dançabilidade não torna a música mais popular.

O Gráfico 4 demonstra de outra forma a correlação.

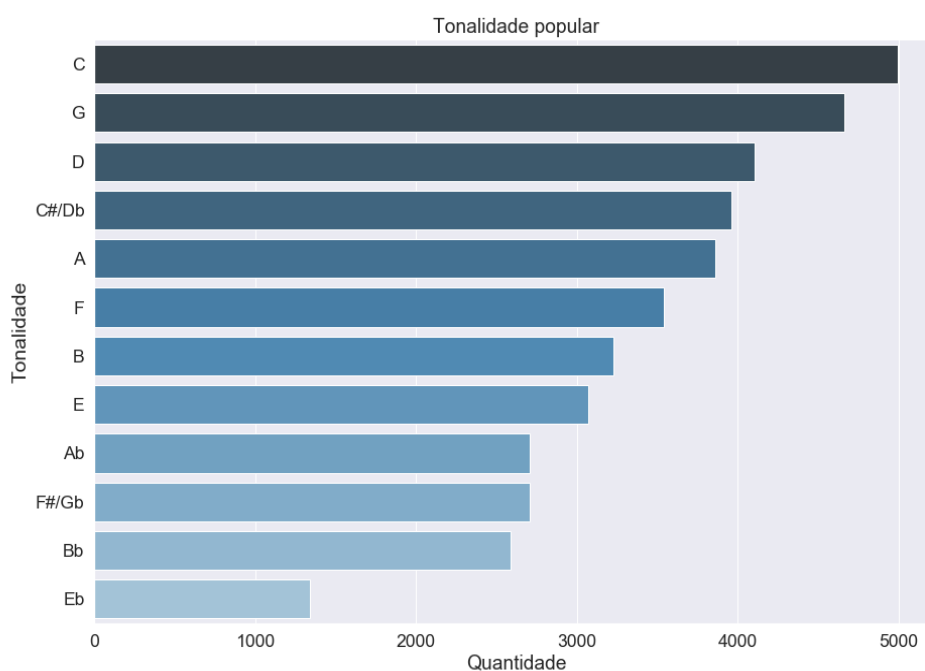
**Gráfico 4** – Outra forma de Correlação



**Fonte:** Elaborado pelo autor, 2019.

O gráfico demonstra de forma mais clara a correlação entre os atributos. A barra à direita indica o nível de correlação, quanto maior o nível, mais escura a cor se torna. Nesse sentido, pode ser visto que a correlação da Dançabilidade, nosso atributo foco, tem maior correlação (0.56) com a positividade, comprovando de tal modo, o que foi visto no Gráfico 3, onde a inclinação da linha vermelha é mais acentuada. Assim como a energia, que possui uma correlação maior com a positividade (0.46). O Gráfico 4, também, comprova que não há uma correlação entre a música estar em uma tonalidade maior ou menor. Temos, portanto, uma visão mais geral das correlações existentes entre os principais atributos estudados.

Sobre a tonalidade, ela reside na ordem em que são dispostos os sons da escala, em suas distâncias respectivas e em suas relações harmônicas (FÉTIS, 1864, *apud* FREITAS; LIBRELOTTO, 2016). No Gráfico 5, pode identificar-se a tonalidade mais presente.

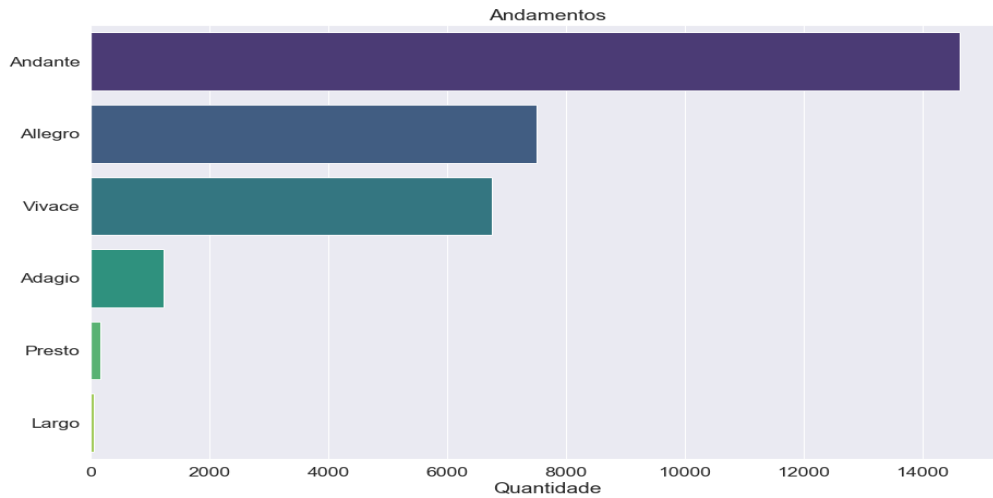
**Gráfico 5** – tonalidade mais presente

**Fonte:** Elaborado pelo autor, 2019.

A tonalidade de C (Dó), é a mais presente, sendo a escala básica de estudos, pois não há alterações, pois, todas as suas notas são as notas naturais, ou seja, sem sustenidos ou bemóis. De certa forma, pode ser que os artistas que possuem músicas mais dançantes, tem a preferência por tonalidades naturais, como mostrado no Gráfico 5. Há sentido nesta afirmação, pois **tonalidades naturais tendem a ser mais agradáveis aos ouvidos.**

A música possui muitas características, dentre elas o tempo. Este tempo está ligado ao ritmo, e assim, ao andamento. Esse é o movimento rápido ou lento dos sons, guardando a precisão dos tempos do compasso (PRIOLLI, 2006). O andamento pode ser caracterizado como lentos, moderados e rápidos. O Gráfico 6, mostra os andamentos mais frequentes encontrados na pesquisa.

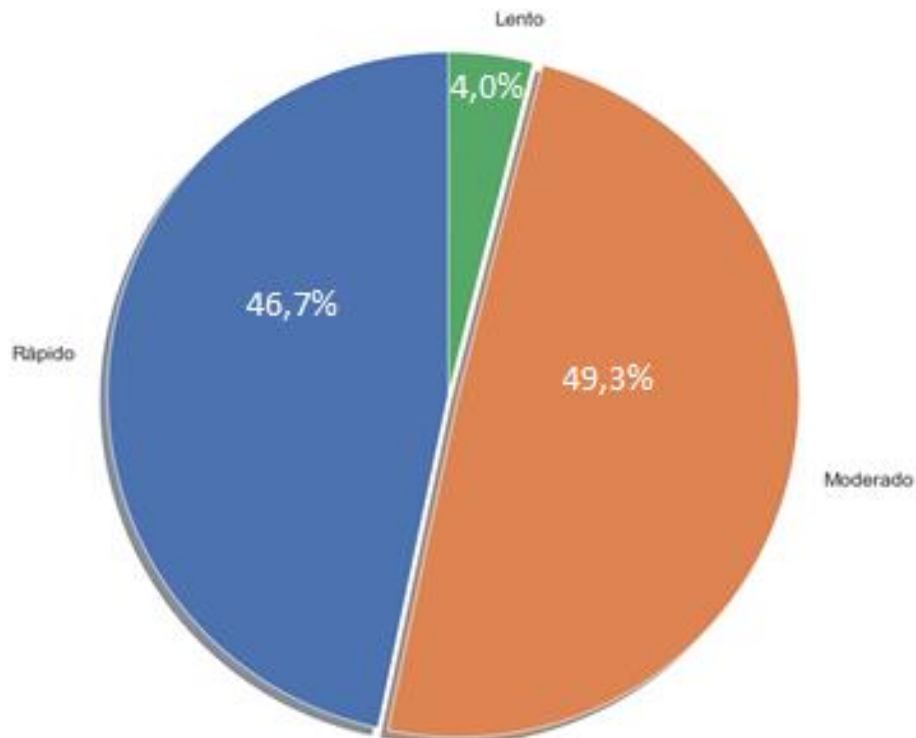
**Gráfico 6 - Andamentos mais frequentes encontrados na pesquisa**



**Fonte:** Elaborado pelo autor, 2019.

No Gráfico 6, é observado o Andante sendo predominante, Pelo gráfico, podemos verificar que a preferência de andamentos, nas músicas pesquisadas, são de moderados a rápidos.

**Gráfico 7- Preferência de andamentos**

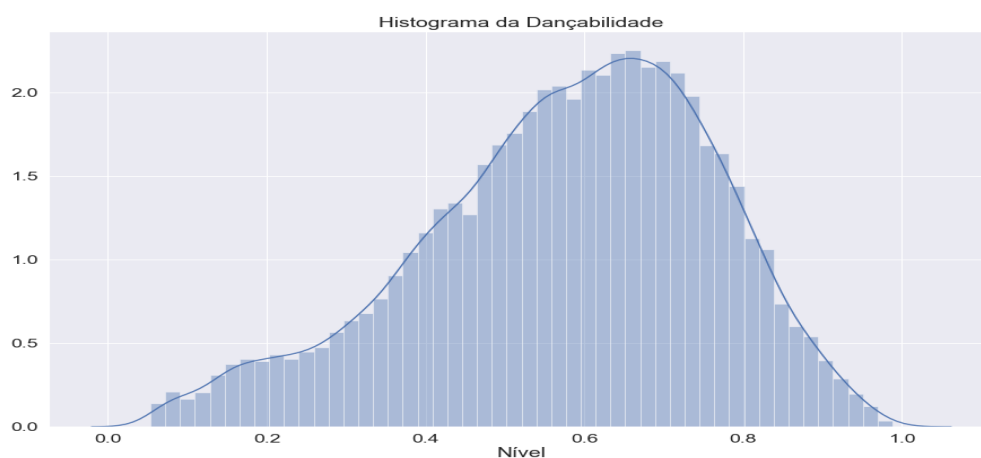


**Fonte:** Elaborado pelo autor, 2019.

Fica claro, que o andamento com predominância nos dados são os Moderados e os Rápidos. Com a preferência um pouco acima para o Moderado.

Com o intuito de verificar o nível de dançabilidade das músicas, o Gráfico 8 mostra o histograma, com a distribuição de frequência. A qual podemos correlacionar, visualmente, que, de fato, um nível de dançabilidade mais alto condiz com andamentos mais rápidos.

**Gráfico 8** – Histograma da Dançabilidade



**Fonte:** Elaborado pelo autor, 2019.

A concentração do Nível de Dançabilidade está entre 0.6 e 0.8. Temos uma Distribuição Assimétrica a Esquerda, ou negativa, pois, a média é menor que a mediana, e essa, menor que a moda. Como mostrado na Tabela 2.

**Tabela 2** – Nível de Dançabilidade

Dançabilidade	
<b>Média</b>	0.57
<b>Mediana</b>	0.60
<b>Moda</b>	0.71

**Fonte:** Elaborada pelo autor, 2019.

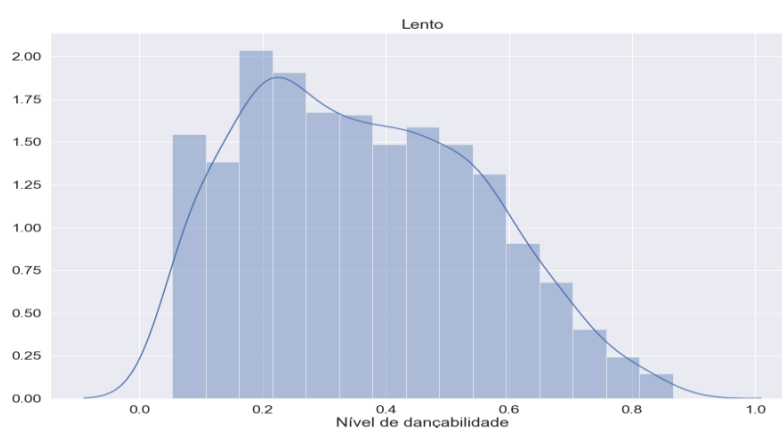
Os dados foram divididos em três grupos de andamento, para ser verificado a distribuição de cada grupo de andamento, com base na Dançabilidade. Os grupos foram divididos conforme Tabela 3.

**Tabela 3** – Grupos de andamento

Andamento	Tempo (bpm)
<b>Lento</b>	Até 75
<b>Moderado</b>	Entre 75 e 120
<b>Rápido</b>	Maior que 120

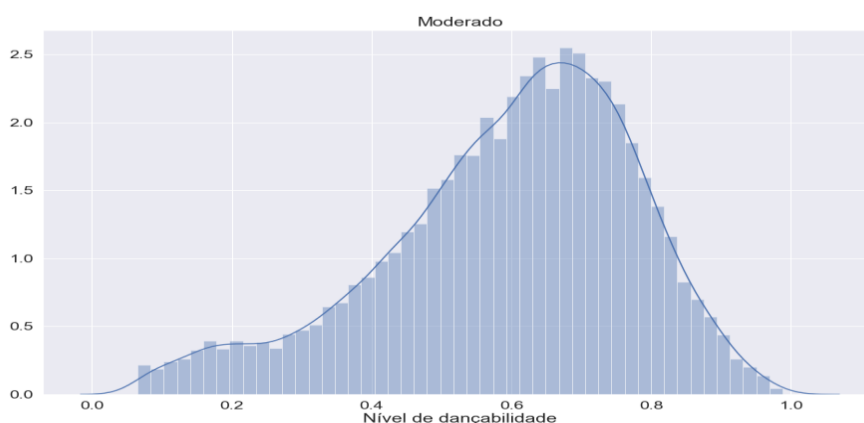
Fonte: Elaborada pelo autor, 2019.

Com base neste agrupamento, os Gráficos 9, 10 e 11, mostram a distribuição da dançabilidade para cada andamento.

**Gráfico 9** – Distribuição da dançabilidade – Andamento Lento

Fonte: Elaborado pelo autor, 2019.

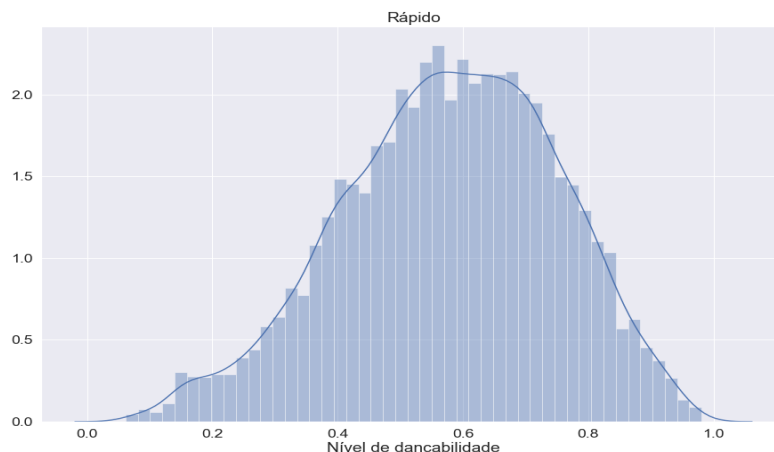
O Gráfico 9, observa-se que o nível de dançabilidade é, de fato, muito baixo, pois se trata de uma dançabilidade muito baixa. A concentração maior de dados fica entre 0.2 e 0.3, o que condiz, com o tempo abaixo de 75bpm, pois se trata de um andamento lento, neste sentido, não é algo tão dançante.

**Gráfico 10** – Distribuição da dançabilidade – Andamento Moderado

Fonte: Elaborado pelo autor, 2019.

O Gráfico 10 mostra uma distribuição mais concentrada entre 0.6 e 0.7, que, de fato, é uma dançabilidade maior.

**Gráfico 11** – Distribuição da dançabilidade – Andamento Rápido

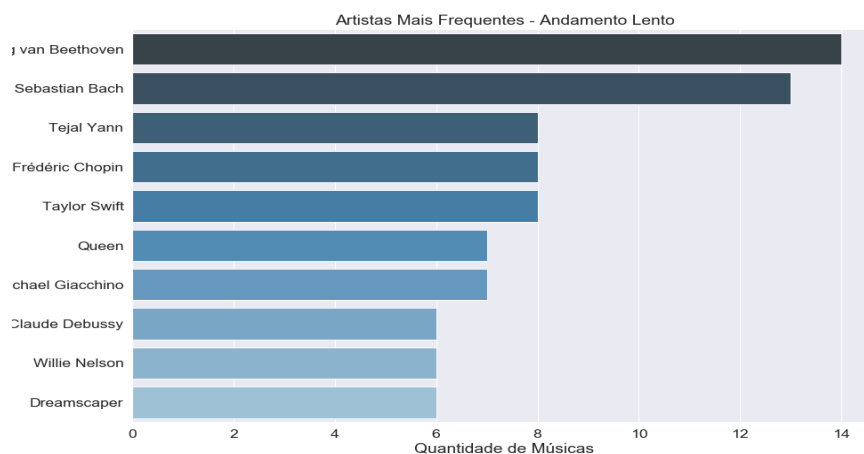


**Fonte:** Elaborado pelo autor, 2019.

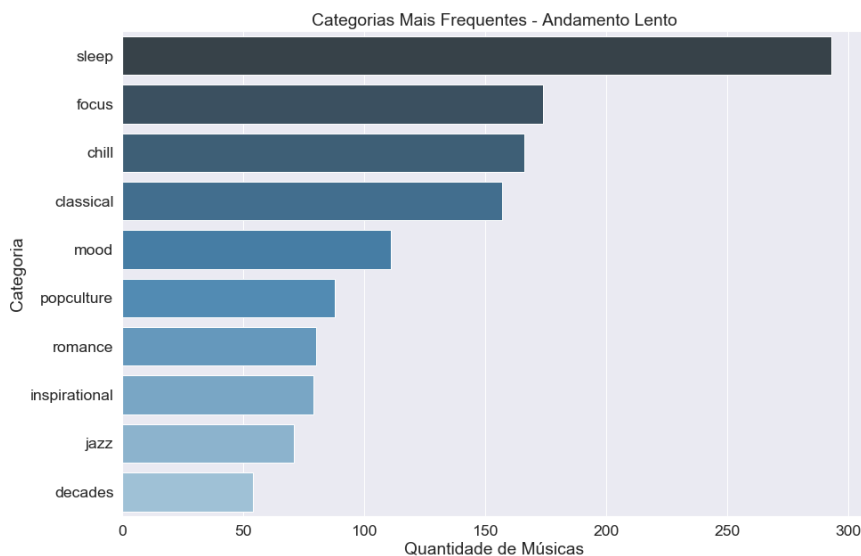
O Gráfico 11, assim como o Gráfico 10, possui uma distribuição com maior concentração em níveis mais altos de dançabilidade. Para o Gráfico 11, o nível de dançabilidade tem uma abrangência um pouco maior, entre 0.5 e 0.8.

Algo interessante a ser levado em consideração, na análise, é saber, quais os artistas mais frequentes em cada grupo de andamento. Os gráficos seguintes têm como objetivo mostrar os artistas mais frequentes, e quais suas características.

**Gráfico 12 (a)** – Artistas mais frequentes – Andamento Lento

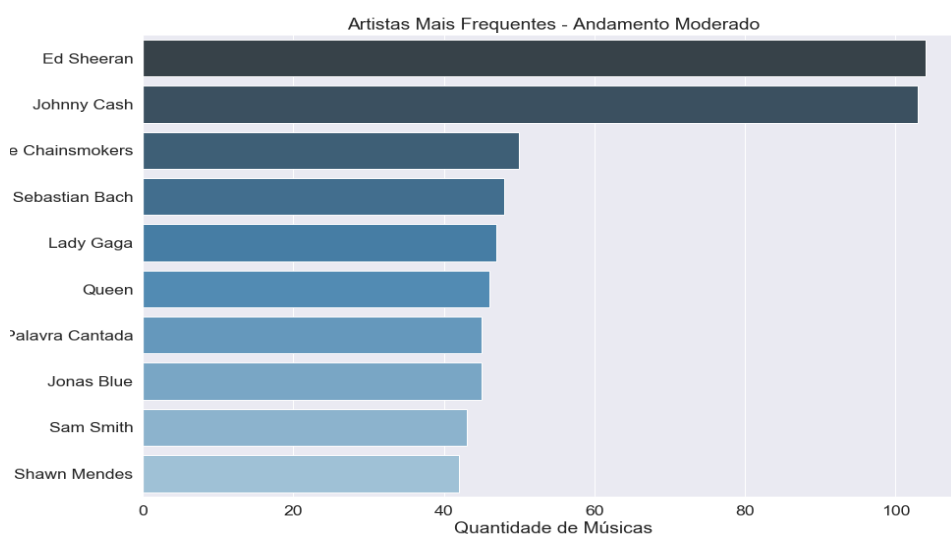


**Fonte:** Elaborado pelo autor, 2019.

**Gráfico 12 (b)** – Categorias mais frequentes – andamento Lento

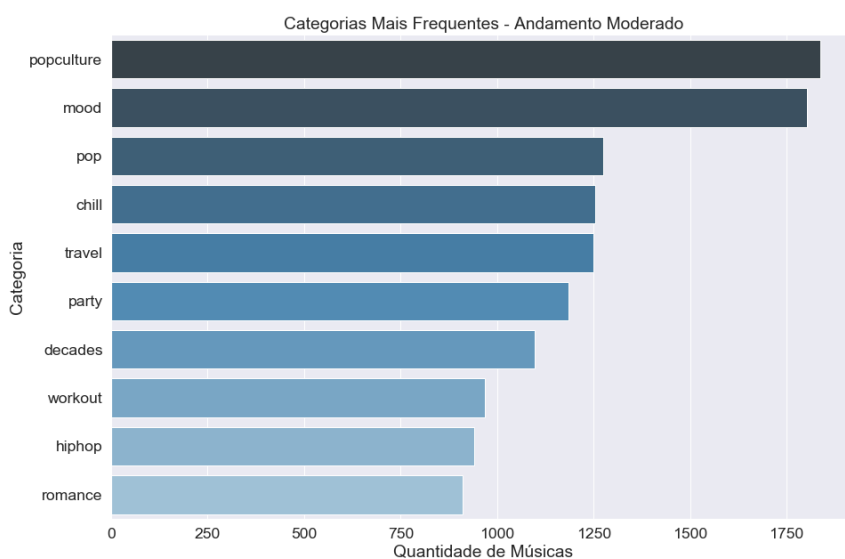
**Fonte:** Elaborado pelo autor, 2019.

Podemos observar no Gráfico 12 (a), a predominância de artistas clássicos, como Beethoven, Bach, Chopin. Artistas estes muito ouvidos, principalmente pelos ouvintes do Japão, Europa e EUA, pela pesquisa, foi constatado que no Brasil esse estilo não é muito escutado, apenas uma pequena fatia da população e estudantes bacharelados em música. Já o Gráfico 12(b), podemos observar as categorias mais frequentes, onde há, realmente, a predominância de categorias que podemos considerar mais lentas, *sleep* (dormir), *focus* (foco).

**Gráfico 13 (a)** – Artistas mais frequentes – Andamento Moderado

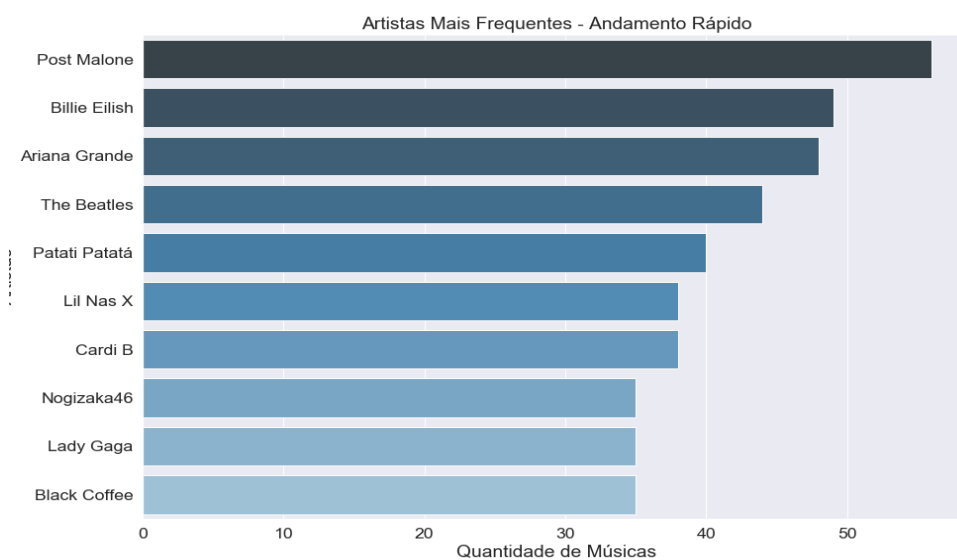
**Fonte:** Elaborado pelo autor, 2019.



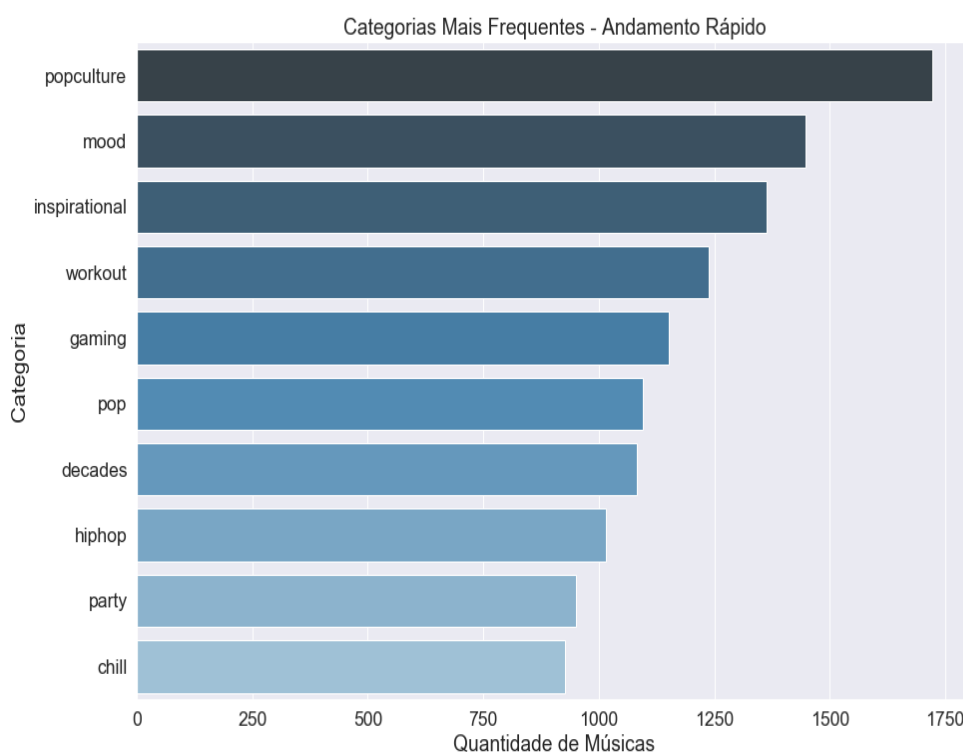
**Gráfico 13 (b)** – Categorias mais frequentes – Andamento Moderado

**Fonte:** Elaborado pelo autor, 2019.

No Gráfico 13(a), é percebido que os artistas são mais variados, contendo desde artistas clássicos, Sebastian Bach; pop, Lady Gaga; rock, Queen; e infantis, Palavra Cantada. No Gráfico 13(b), percebe a predominância da cultura pop e mood.

**Gráfico 14 (a)** – Artistas mais frequentes – Andamento rápido

**Fonte:** Elaborado pelo autor, 2019.

**Gráfico 14 (b)** – Categorias mais frequentes – Andamento Rápido

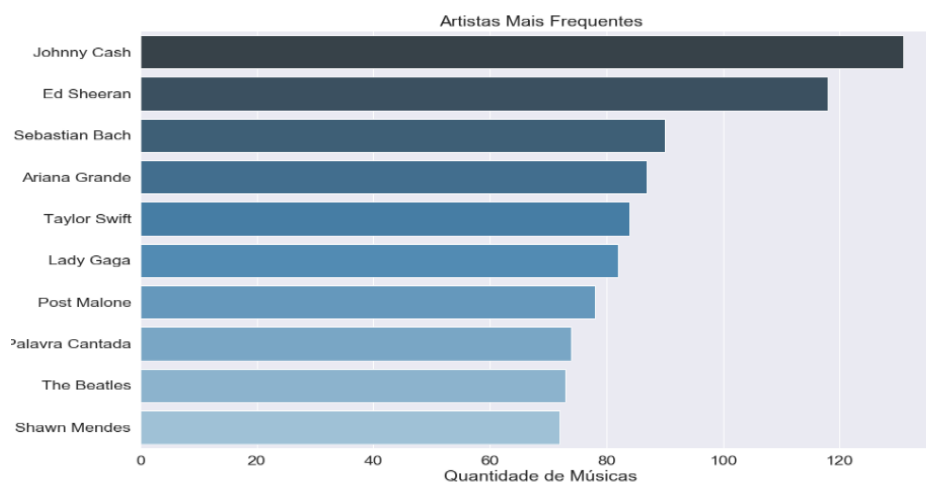
**Fonte:** Elaborado pelo autor, 2019.

No Gráfico 14(a), percebe-se a predominância de artistas pops, contendo ainda artistas infantis. O Gráfico 14(b), confirma a predominância da cultura pop nas músicas com andamento rápido. Além de músicas que remetem **energia**, atributo que tem uma alta correlação com a **dançabilidade**, como: *inspirational* (inspiração), *workout* (ouvido muito em academias).

#### 4.1.2 Análise por artistas

Nesta subseção, o objetivo é identificar os artistas mais frequentes, a partir desses, os três mais frequentes serão analisados, de forma individual. Identificando as características de dançabilidade.

### Gráfico 15 – Artistas mais frequentes

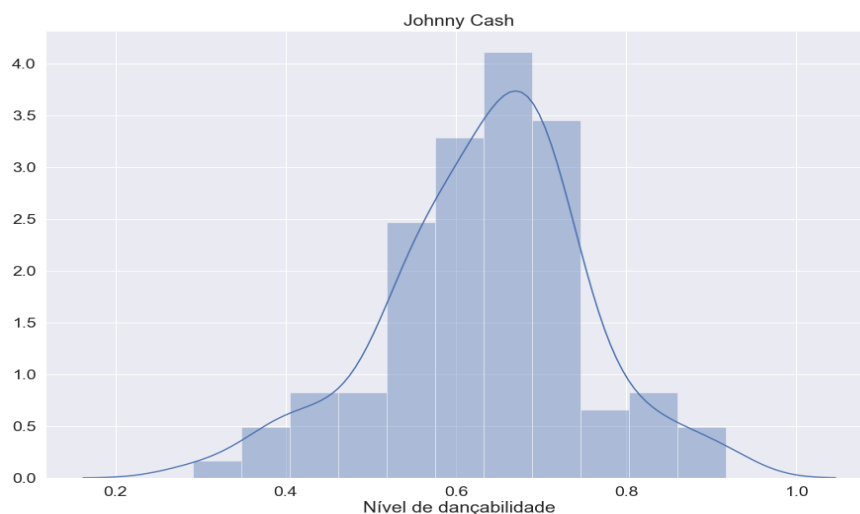


**Fonte:** Elaborado pelo autor, 2019.

O Gráfico 15, mostra os artistas mais frequentes. Os artistas com maior predominância de músicas são: Johnny Cash, Ed Sheeran e Johann Sebastian Bach. Há uma variedade de estilos entre os países, categorias com características distintas, como o country, pop e clássico.

#### 4.1.2.1 Johnny Cash

Mais conhecido como **O homem de preto**, foi um cantor, compositor, escritor, diretor e ator norte-americano. Considerado um dos músicos mais influentes do século XX, é referência no gênero country (WIKIPÉDIA, 2019). Estilo predominante na categoria do *Spotify*. As músicas mais tocadas do artista são: *I Walk the Line* e *Ring of Fire*.

**Gráfico 16** – Artista Johnny Cash – Nível de dançabilidade

**Fonte:** Elaborado pelo autor, 2019.

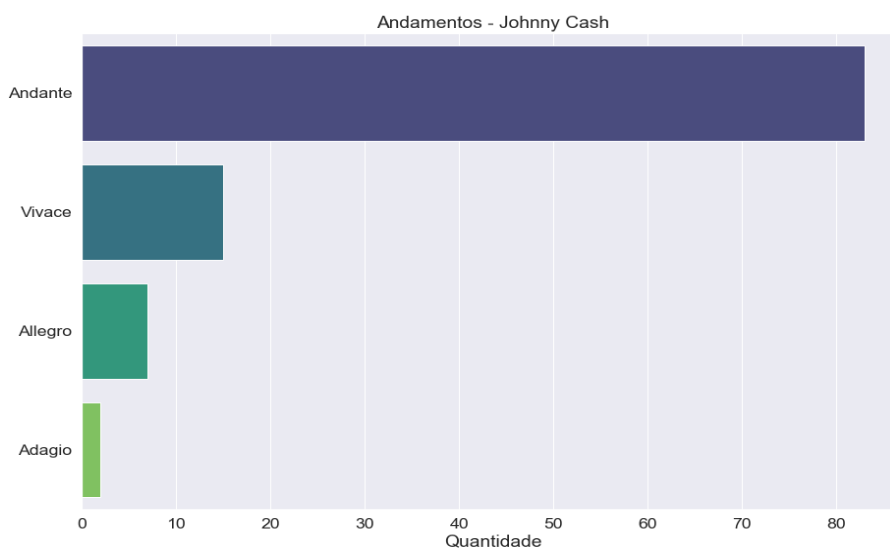
O nível de dançabilidade das suas músicas é uma distribuição normal, conforme Gráfico 16 centralizada em 0.65. Mostrado na Tabela 4.

**Tabela 4** – Nível de Dançabilidade

Dançabilidade	
<b>Média</b>	0.64
<b>Mediana</b>	0.65
<b>Moda</b>	0.65

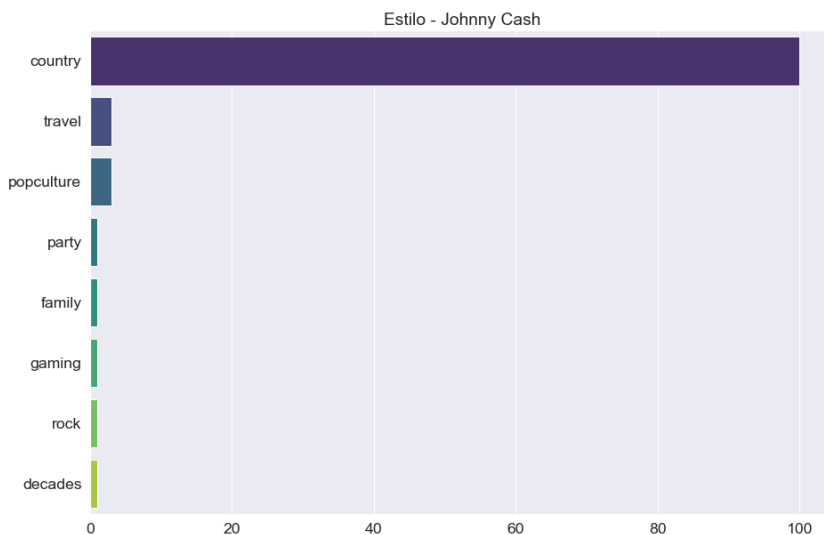
**Fonte:** Elaborada pelo autor, 2019.

O estilo musical de Johnny Cash pode ser considerado dançante, e se encaixa em um andamento moderado, com a maioria de suas músicas em andamentos mais rápidos. Mostrado no Gráfico 17.

**Gráfico 17 – Andamentos – Johnny Cash**

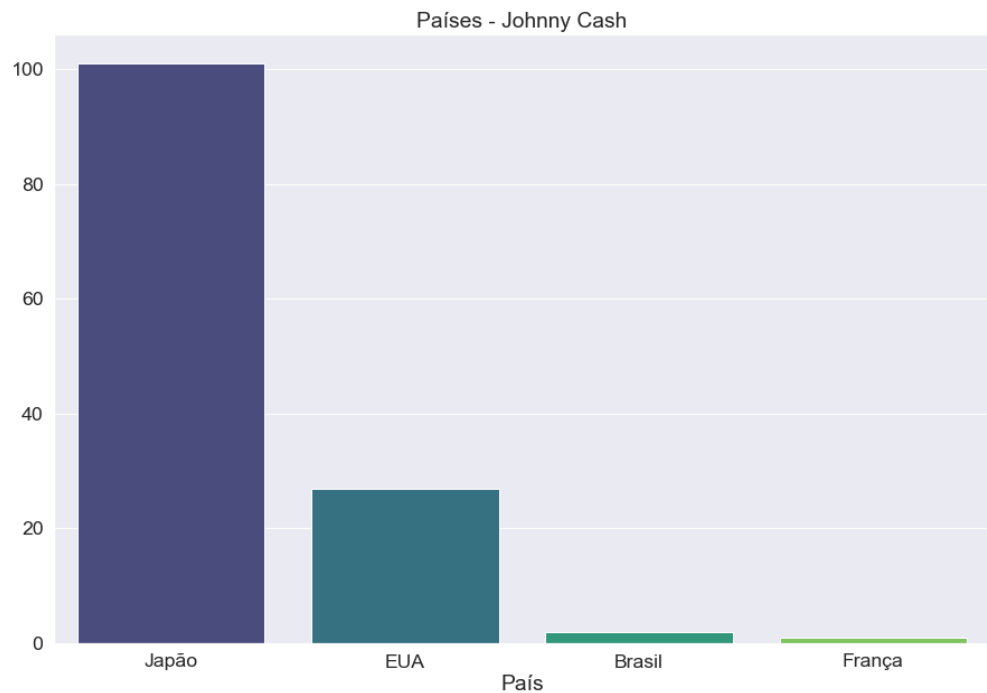
**Fonte:** Elaborado pelo autor, 2019.

A categoria musical do cantor é bem eclética, passando por vários estilos. Conforme Gráfico 18, a predominância é o country.

**Gráfico 18 – Estilo Musical – Johnny Cash**

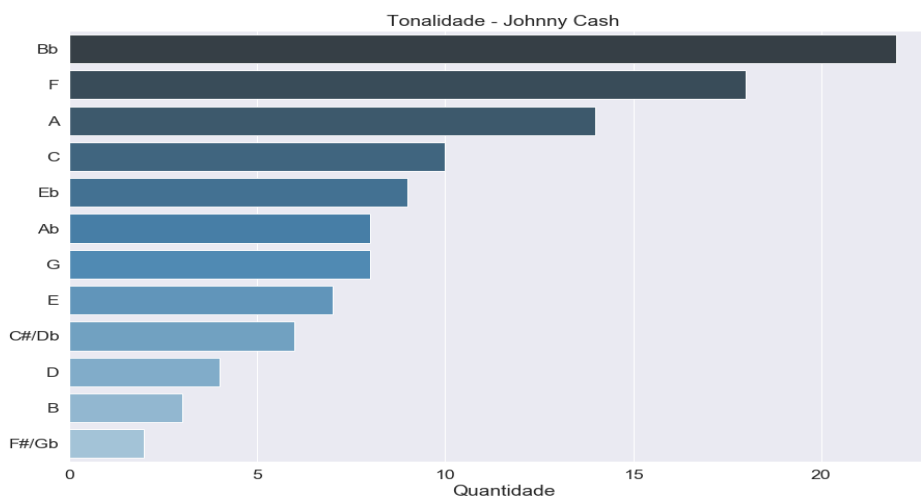
**Fonte:** Elaborado pelo autor, 2019.

Apesar do seu estilo ter um foco na música *country*, que é um estilo de origem americana, as músicas de Cash são ouvidas, predominantemente, em outros países.

**Gráfico 19** – Países onde são ouvidas as músicas de Johnny Cash

**Fonte:** Elaborado pelo autor, 2019.

O Gráfico 19 aponta para o Japão como o país que mais escuta as músicas de Johnny Cash através do *Spotify*. Os Estados Unidos apresentam uma certa quantidade de músicas, que ainda pode ser relevante, diferentemente do Brasil e França, onde poucas músicas chegam a ser escutadas. O Gráfico 20 mostra as características de tonalidade do artista.

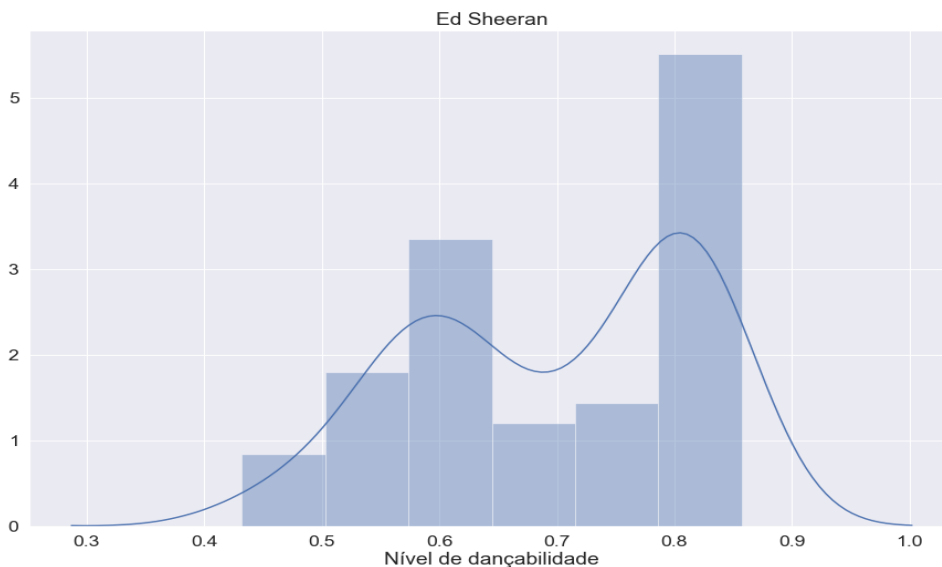
**Gráfico 20** – Características de Tonalidade – Johnny Cash

**Fonte:** Elaborado pelo autor, 2019.

A tonalidade mais frequente das músicas de Johnny Cash é a de Bb (Si Bemol), F (Fá) e A (Lá).

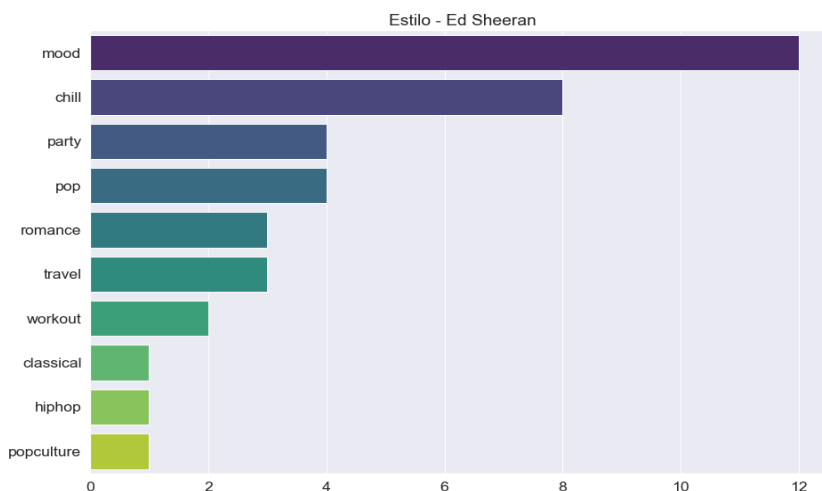
#### 4.1.2.2 Ed Sheeran

Edward Christopher, “Ed Sheeran”, cantor e compositor britânico, ficou bastante conhecido a partir de 2012, principalmente nos Estados Unidos, após participação no quarto álbum de estúdio de Taylor Swift. Sheeran começou a cantar com 4 anos, no coral da igreja local, aprendeu a tocar guitarra e escreveu suas próprias músicas quando ainda estava no colegial. Algumas de suas influências musicais são: Damien Rice, The Beatles, Dylan, Nizlopi e Eminem. Seu estilo é voltado ao estilo pop (VAGALUME, 2019). Tendo com músicas mais conhecidas *Shape of You*, *Thinking Out Loud*.

**Gráfico 21 – Ed Sheeran – Nível de Dançabilidade**

**Fonte:** Elaborado pelo autor, 2019.

Conforme Gráfico 21, observamos que seu nível de dançabilidade é uma distribuição bimodal, onde temos dois níveis bem característico em suas músicas que são 0.6 e 0.8. É um artista com características bem dançantes em suas músicas. Suas categorias no *Spotify* são bem diversificadas, no entanto, em sua maioria, consideradas dançantes.

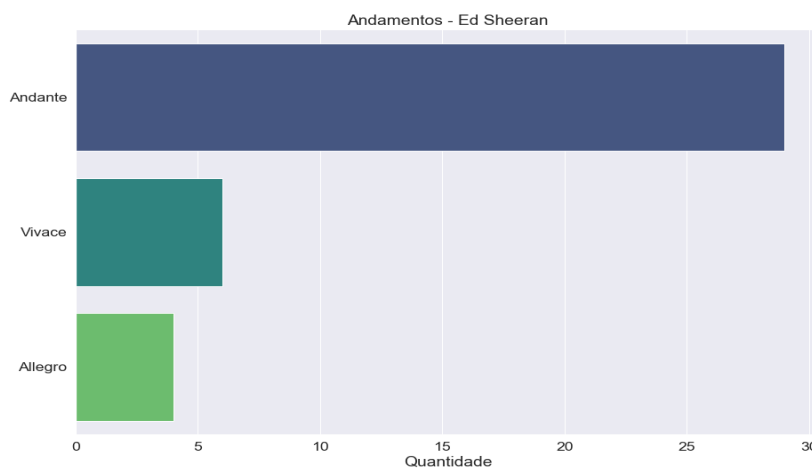
**Gráfico 22 – Estilo musical do artista Ed Sheeran**

**Fonte:** Elaborado pelo autor, 2019.



O Gráfico 22 demonstra bem as características de estilo do artista. Passando pelo *pop*, *popculture*, *travel*, *mood*. Seus estilos dançantes são confirmados pelos andamentos em suas músicas, que são andamentos, conforme Gráfico 22.

**Gráfico 23** – Andamentos nas músicas de Ed Sheeran

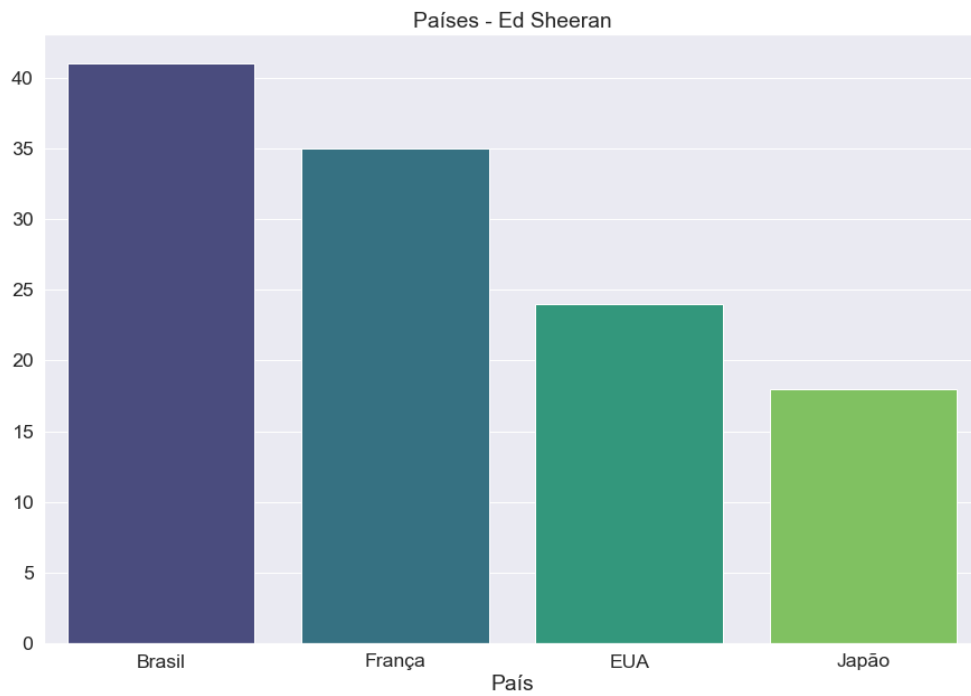


**Fonte:** Elaborado pelo autor, 2019.

As músicas de Ed Sheeran têm uma predominância com um andamento moderado, o andante. Os andamentos que estão presentes em suas músicas são andamentos rápidos, *vivace* e *allegro*, conforme Gráfico 23.

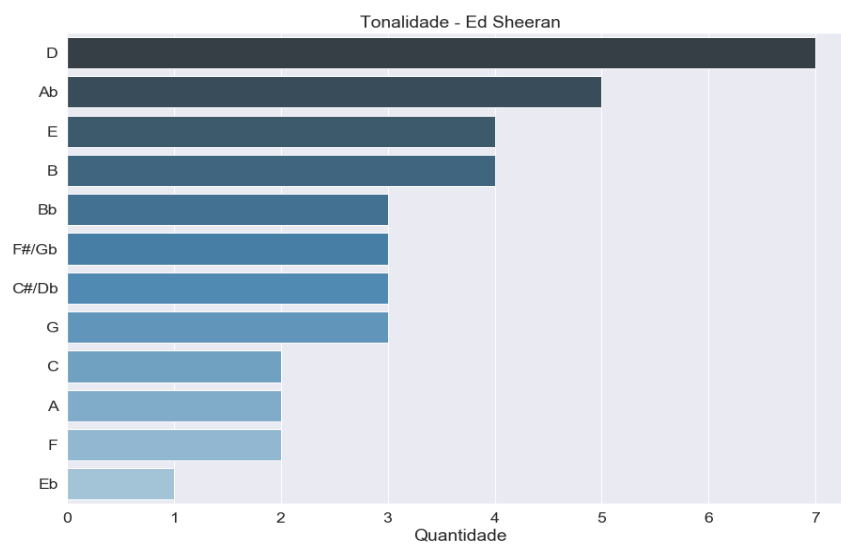
Através do Gráfico 24, é percebido que Ed Sheeran é um artista escutado em vários países, dos analisados, com maior frequência no Brasil. No entanto, está presente em todos os outros países, sendo menos escutado no Japão.

**Gráfico 24 – Países onde Ed Sheeran é mais escutado**



**Fonte:** Elaborado pelo autor, 2019.

**Gráfico 25 – Tonalidade – Ed Sheeran**



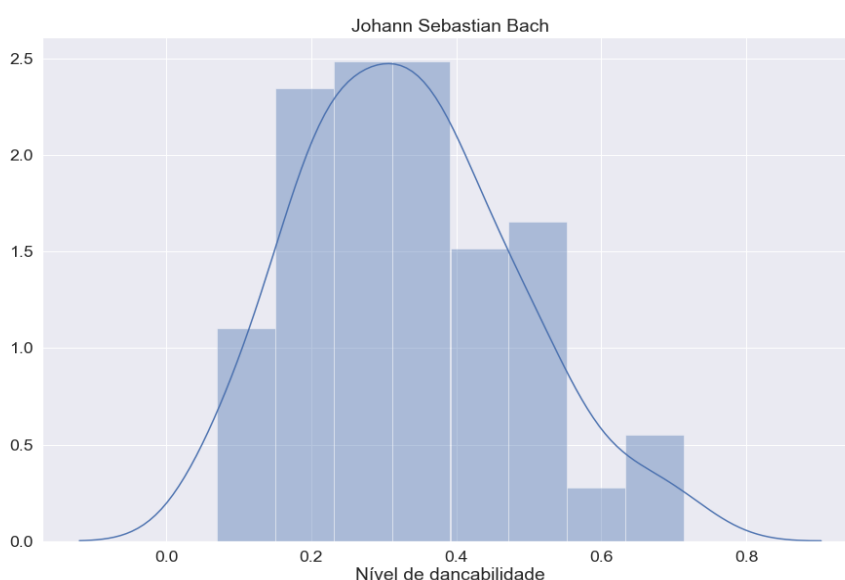
**Fonte:** Elaborado pelo autor, 2019.

No Gráfico 25, podemos perceber as tonalidades predominantes do artista Ed Sheeran, que são o D (Ré), Ab (Lá bemol), E (Mi) e B (Si).

#### 4.1.2.3 Johann Sebastian Bach

Johann Sebastian Bach, compositor, cravista, Kapellmeister, regente, organista, professor, violonista e violista, foi um artista alemão, nasceu em 21 de março de 1685, em Eisenach. É tido como o maior nome da música barroca, além disso, tinha uma obstinação: “combinar melodias da música e, graças a sua técnica, dominou como ninguém a ciência da composição em prol da harmonia perfeita”. Foi estudado por grandes compositores como Mozart e Beethoven (FOLHA ONLINE, 2019). As músicas mais ouvidas, com base nos dados, são: *Toccatà and Fugue in D Minor*, *Orchestral Suite No. 3 in D Major* e *Prelude & Fughetta in G Major*.

**Gráfico 26** – Johann Sebastian Bach – Nível de Dançabilidade



**Fonte:** Elaborado pelo autor, 2019.

Conforme Gráfico 26, a distribuição de seu nível de dançaabilidade é uma distribuição normal. Conforme a Tabela 5 pode ser verificado as características em torno da média, mediana e moda.

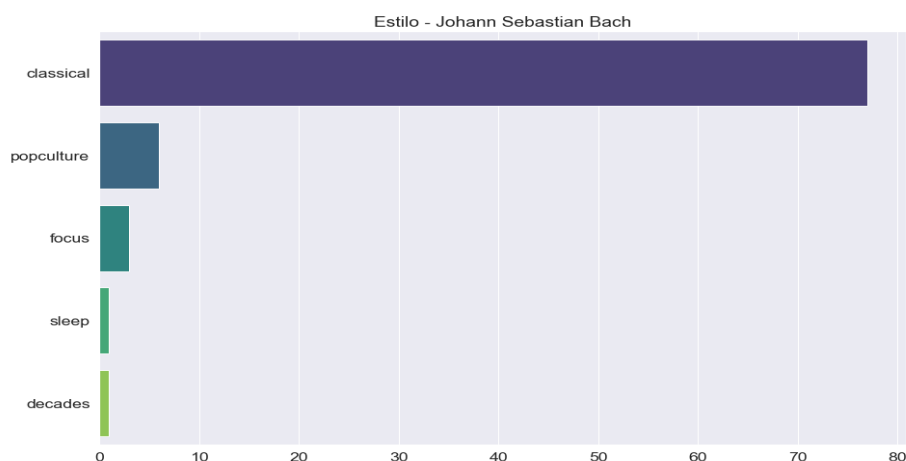
**Tabela 5** – Características do nível de Dançabilidade

Dançabilidade	
<b>Média</b>	0.32
<b>Mediana</b>	0.33
<b>Moda</b>	0.36

**Fonte:** Elaborada pelo autor, 2019.

O nível de dançabilidade de Sebastian Bach não é dançante. Isso pelo seu estilo, pois é uma música muito mais para ser apreciada do que dançada. Porém, existem obras do artista que possui alto nível de dançabilidade.

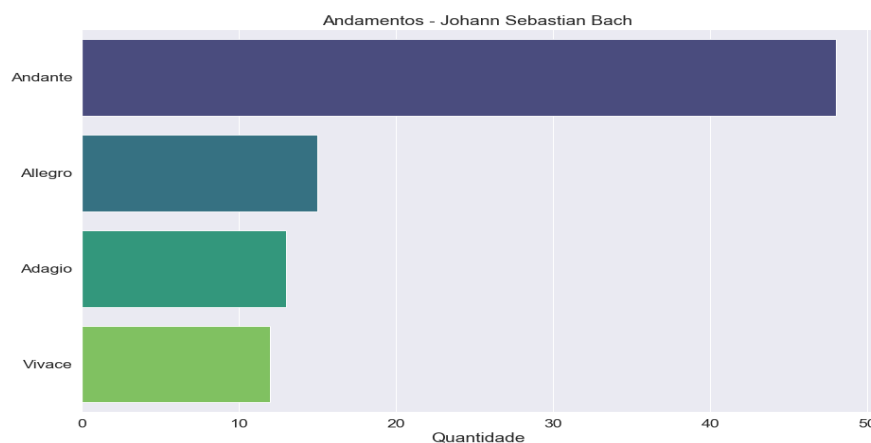
**Gráfico 27** – Estilo musical – Johann Sebastian Bach



**Fonte:** Elaborado pelo autor, 2019.

Conforme Gráfico 27, a categoria principal de Bach é o clássico, o que é, de fato, o estilo musical do artista. Podendo ser categorizado em outros estilos. Quanto ao andamento de suas músicas, o Gráfico 28 mostra predominância pelo andante, sendo um andamento moderado. Possui músicas em outros andamentos mais rápidos: allegro e vivace. E algumas obras no andamento lento: adagio. Conforme Gráfico 28.

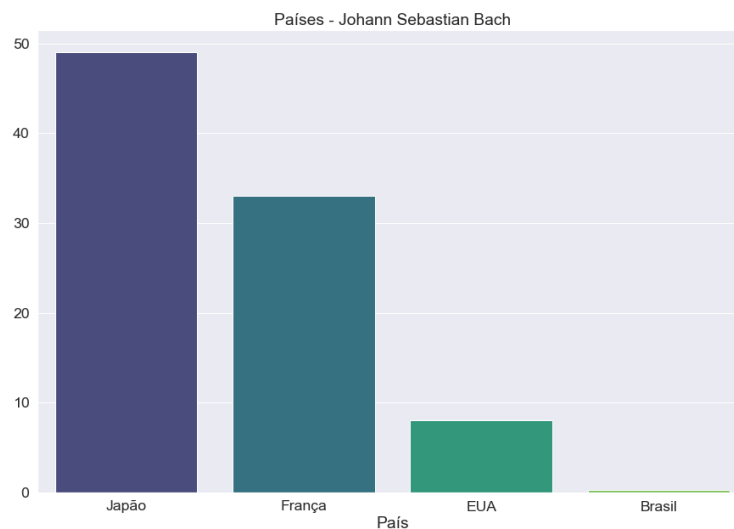
**Gráfico 28** – Andamentos – Johann Sebastian Bach



**Fonte:** Elaborado pelo autor, 2019.

O Gráfico 29 mostra os países que possuem mais músicas do artista Bach.

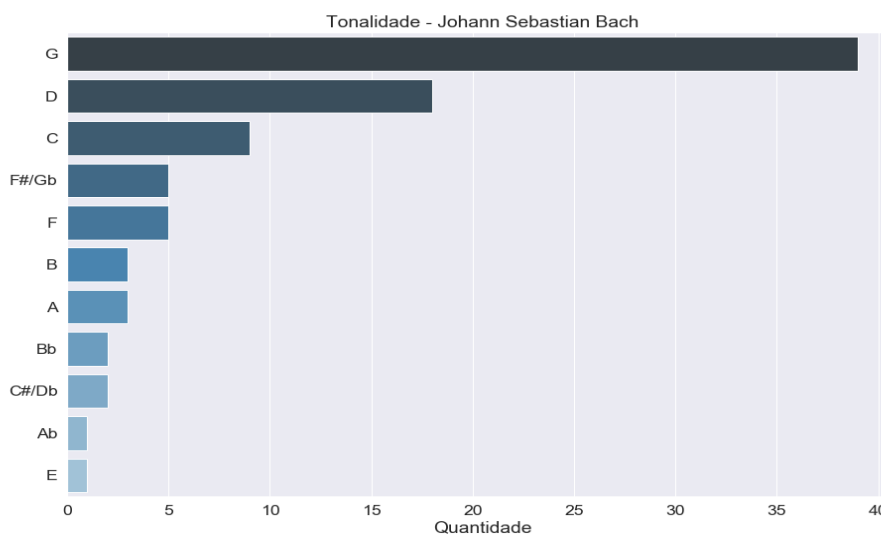
**Gráfico 29** – Países onde mais se escuta Johann Sebastian Bach



**Fonte:** Elaborado pelo autor, 2019.

O Japão, Bach é o país que apresenta a maior concentração de músicas do artista. Seguido da França. O Brasil não possui músicas do artista como mais populares. Conforme Gráfico 30, a tonalidade das obras do artista tem como maioria a tonalidade de G (Sol).

**Gráfico 30** – Tonalidade – Johann Sebastian Bach

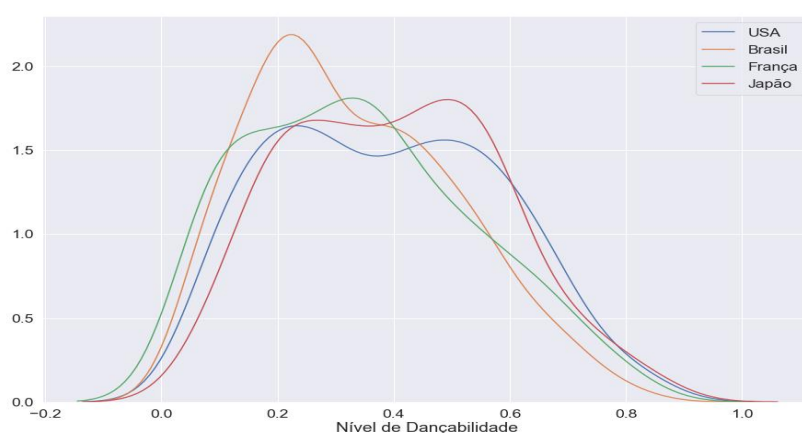


**Fonte:** Elaborado pelo autor, 2019.

### 4.1.3 Análise de Dados das músicas por País

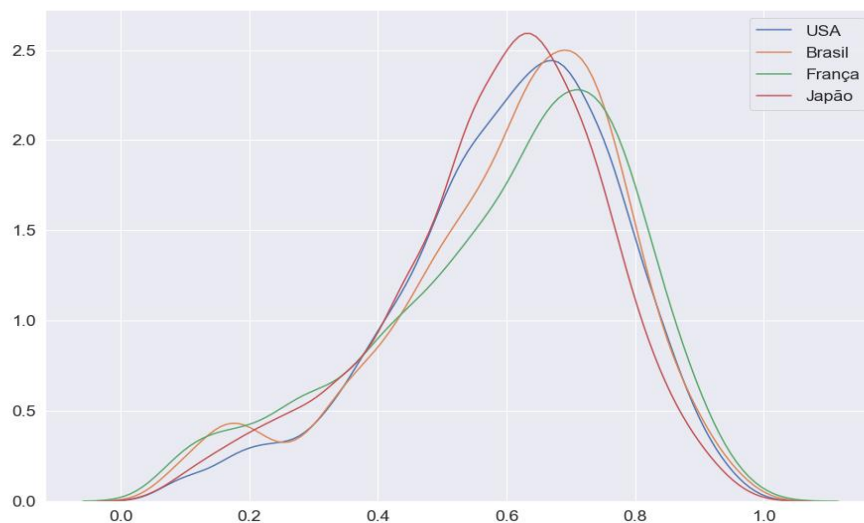
Nesta parte, após ser analisado os dados de forma geral, vamos ter, no entanto, uma visualização, ainda geral, verificando os níveis de dançabilidade por país, comparando com cada um dos três grupos de andamento abordados: lento, moderado e rápido.

**Gráfico 31** – Nível de Dançabilidade por país – Andamento Lento



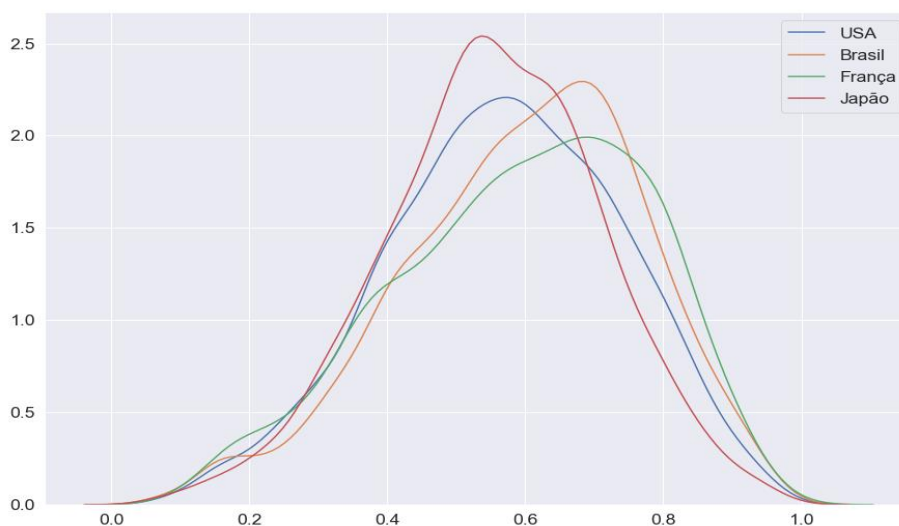
**Fonte:** Elaborado pelo autor, 2019.

No Gráfico 31, podemos perceber níveis de dançabilidade com base no andamento lento. Cada país possui sua característica nas músicas ouvidas, mesmo que sejam poucas. Percebe-se, a priori, que, o Brasil tem uma média maior no nível 0.2. Já no nível 0.4, há um equilíbrio entre os países. Neste grupo, há dançabilidades com um nível mais alto, entre 0.4 e 0.6, o qual tem como país predominante o Japão.

**Gráfico 32 – Níveis de Dançabilidade por país – Andamento Moderado**

**Fonte:** Elaborado pelo autor, 2019.

O Gráfico 32, os níveis de dançabilidade são mais altos, pois esses são do grupo do andamento moderado. Pode ser percebido que o Japão tem uma distribuição maior de dançabilidade no nível 0.6. Já o Brasil, 0.7. A França, neste gráfico, é o que apresenta a maior frequência no nível de dançabilidade acima de 0.8.

**Gráfico 33 – Níveis de Dançabilidade – Andamento Rápido**

**Fonte:** Elaborado pelo autor, 2019.

O Gráfico 33, percebe-se que o Brasil e o Japão são os que apresentam maiores frequências nos níveis entre 0.4 a 0.8. Já a partir do nível 0.8, segue o maior nível para a França.

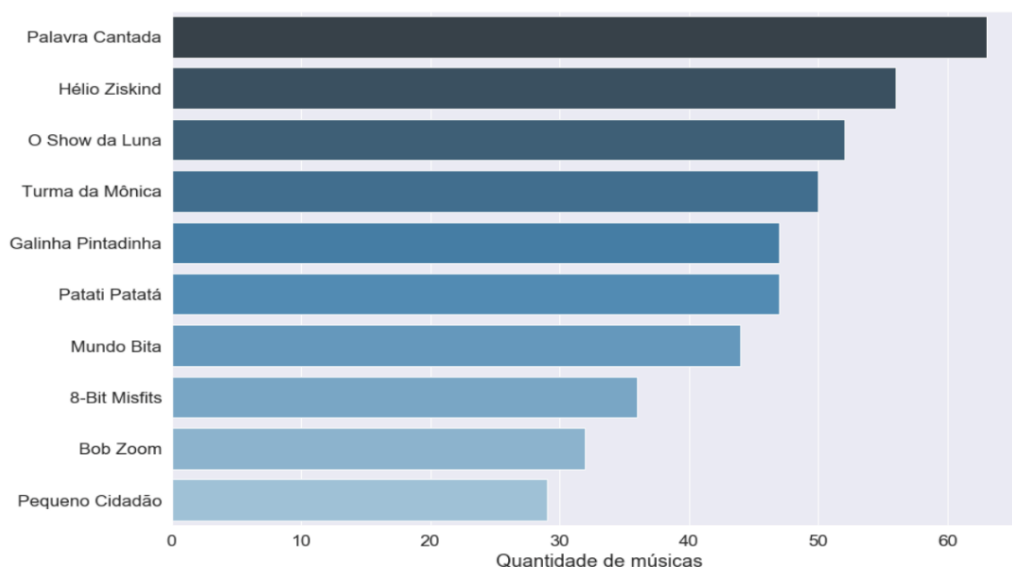
Verificando os gráficos, percebe que o Japão tem uma predominância de dançabilidade entre 0.4 e 0.7. Já o Brasil, é 0.2, e 0.7. No caso da França, sua predominância é nos níveis acima de 0.8. Já os Estados Unidos até 0.6, apresenta maior preferência do que Brasil e França

Podemos, a priori, afirmar que a França é o país que apresenta escolhas de músicas com um maior nível de dançabilidade. Já o Japão, mantém, em todos os grupos, uma média quase que constante.

#### 4.1.3.1 Brasil

O Gráfico 34 apresenta os artistas mais comuns para os dados do Brasil.

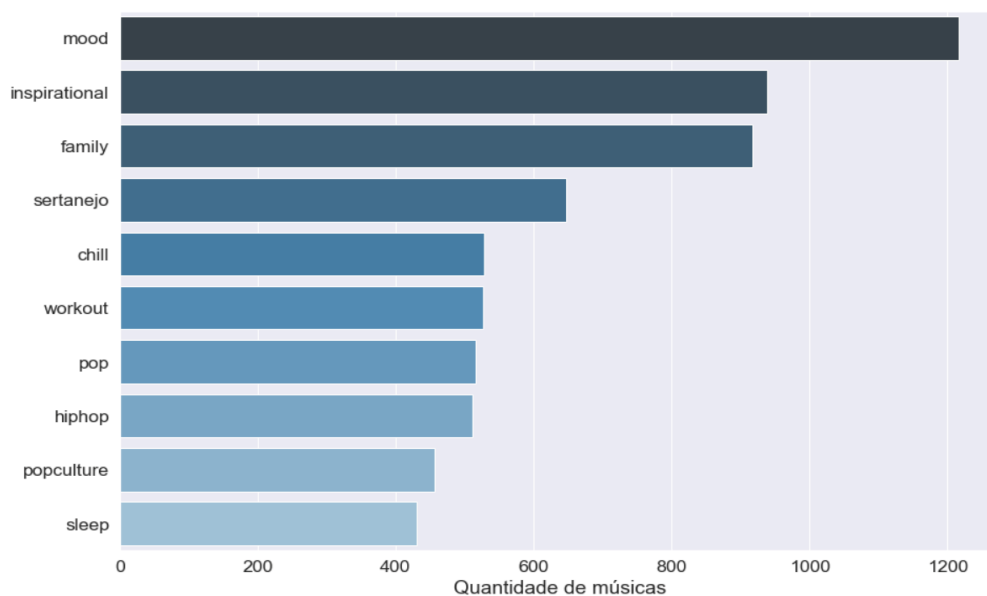
**Gráfico 34** – Artistas mais comuns no Brasil



**Fonte:** Elaborado pelo autor, 2019.

Há a predominância de artistas infantis, como por exemplo, Palavra Cantada, Galinha Pintadinha, O Show da Luna, Turma da Mônica entre outros, o que pode denotar que os pais das crianças optam por mídias como o *Spotify* para entretê-las. O Gráfico 35 mostra as categorias mais comuns no país.



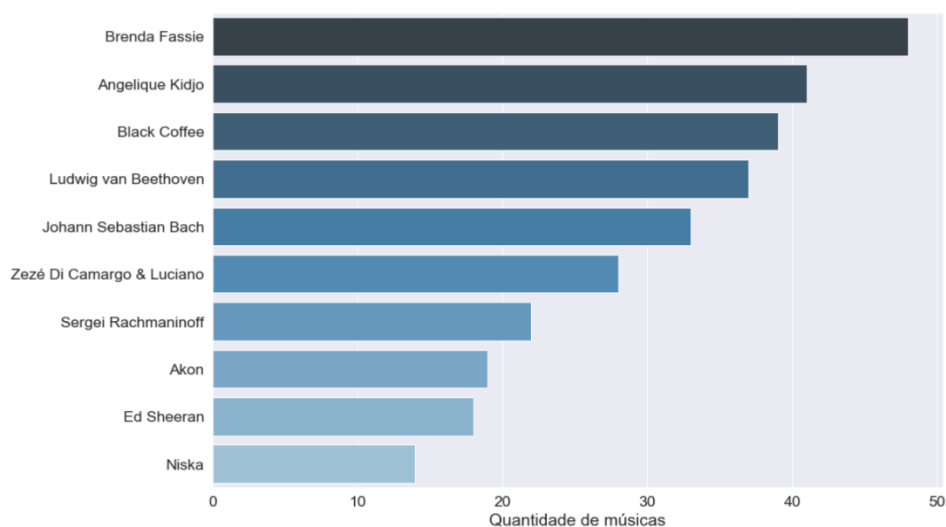
**Gráfico 35** – Categorias musicais mais comuns no Brasil

**Fonte:** Elaborado pelo autor, 2019.

O estilo mais comum é o *mood*, porém, apesar disso, os artistas mais frequentes são artistas infantis, o qual dependendo das características, podem estar presentes em outras categorias do *Spotify*.

#### 4.1.3.2 França

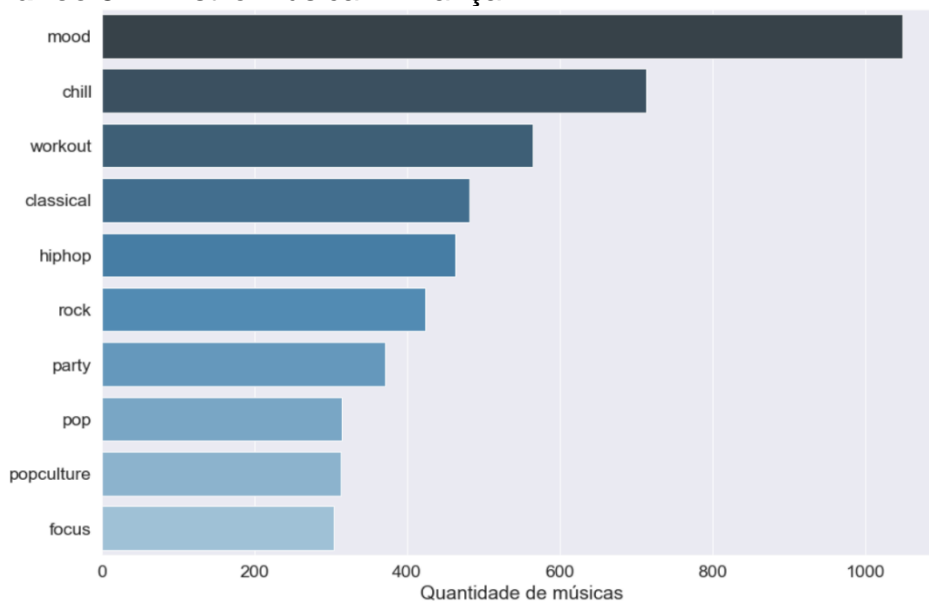
O Gráfico 36 apresenta os artistas mais comuns para os dados da França.

**Gráfico 36** – Artistas que mais ouvidos na França

**Fonte:** Elaborado pelo autor, 2019.

A França, conforme Gráfico 36, possui artistas de vários estilos e países, incluindo artistas brasileiros, como Zezé Di Camargo & Luciano. Sendo Brenda Fassie, uma artista pop sul-africana, com maior número de músicas encontradas.

**Gráfico 37 – Estilo Musical – França**

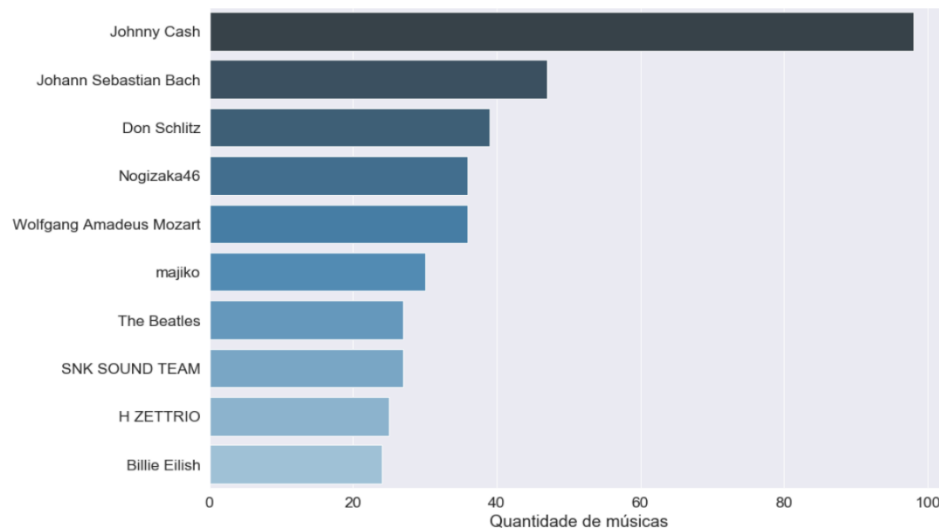


**Fonte:** Elaborado pelo autor, 2019.

Conforme Gráfico 37, o *mood*, assim como no Brasil, é bem mais frequente, no entanto, contém vários outros estilos como o *classical*, *hip-hop*, *popculture*.

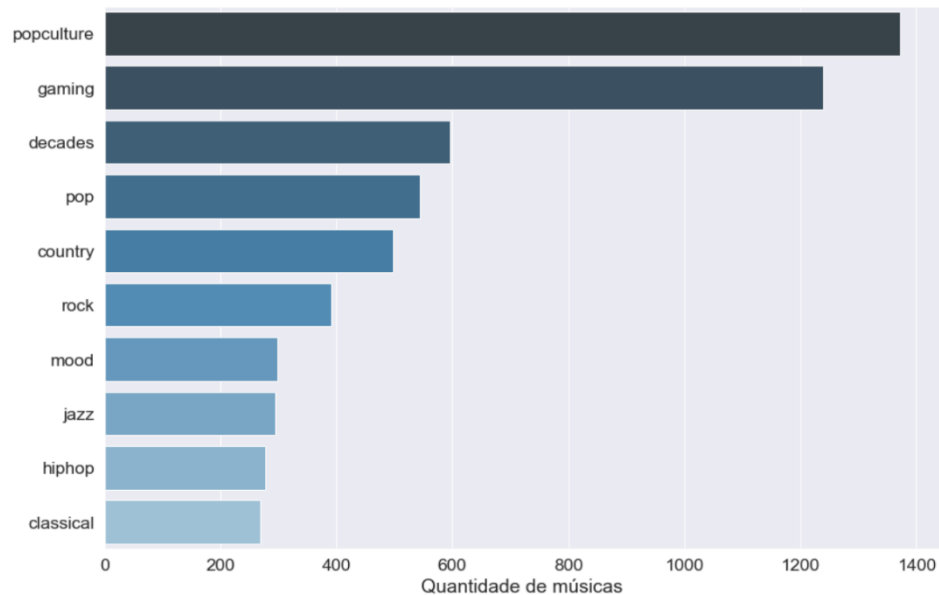
#### 4.1.3.3 Japão

O Japão possui um gosto musical mais voltado à clássicos, de décadas, por apresentar artistas como Johnny Cash, Bach, The Beatles. Conforme Gráfico 38.

**Gráfico 38** – Artistas mais ouvidos no Japão

**Fonte:** Elaborado pelo autor, 2019.

Os estilos encontrados, no país, variam de músicas pop a *jazz*, do clássico ao *hip-hop*. Conforme Gráfico 39:

**Gráfico 39** – Estilo musical - Japão

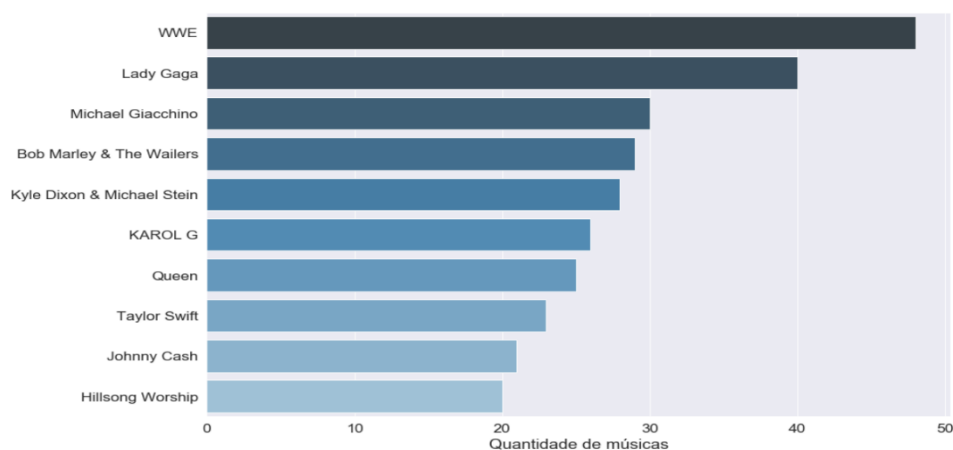
**Fonte:** Elaborado pelo autor, 2019.

#### 4.1.3.4 Estados Unidos

Os Estados Unidos possuem uma preferência mais voltada aos artistas ingleses. Como Lady Gaga, Taylor Swift, Queen. Diferentemente dos outros países,

que podem ser encontrados artistas não apenas dos seus países. Conforme Gráfico 40:

**Gráfico 40** – Artistas ouvidos nos Estados Unidos

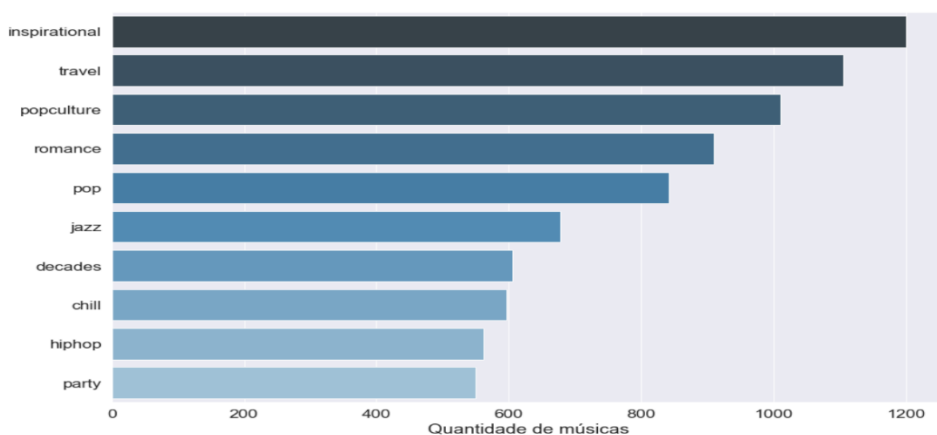


**Fonte:** Elaborado pelo autor, 2019.

A WWE é uma empresa de entretenimento, conhecida pela luta livre profissional. Também realiza trabalhos musicais, e é o artista do *Spotify*, para o país, com maior frequência.

As categorias mais frequentes no país são o *pop*, *jazz*, *hip-hop*. Que são estilos presentes no país. Conforme Gráfico 41:

**Gráfico 41** – Estilos musicais mais frequentes nos EUA

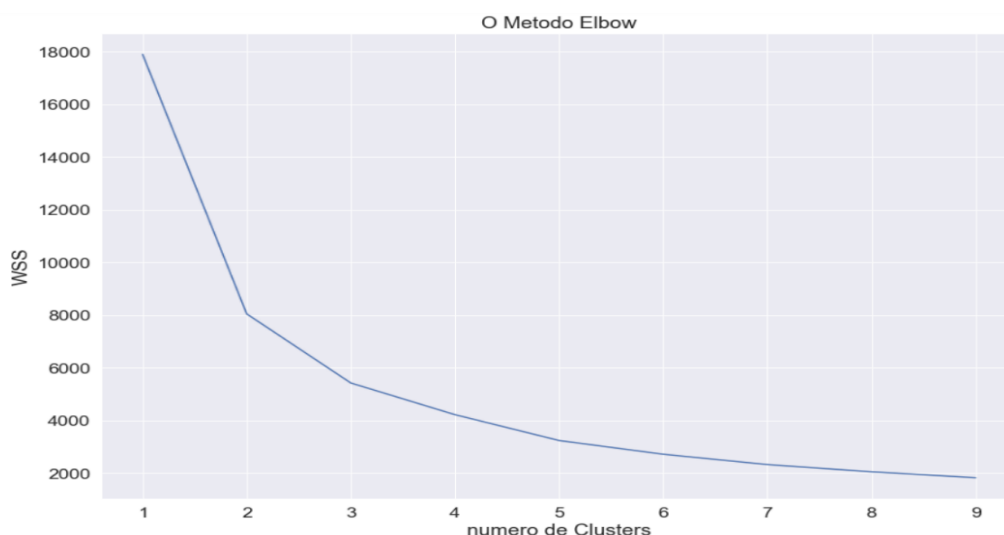


**Fonte:** Elaborado pelo autor, 2019.

## 4.2 Clusterização: quantidade de cluster através do método *Elbow*

A clusterização teve como objetivo agrupar as músicas, independentemente do país, por nível de dançabilidade e positividade. Por serem os atributos com maior correlação, pôde ser mostrado, de forma clara, o agrupamento das músicas através dos níveis. Podemos então perceber que uma música positiva, também é, uma música dançante. A escolha da quantidade de cluster foi feita através do método *Elbow*, que calcula a quantidade de cluster para determinados atributos. Foram utilizados dois atributos para a clusterização, a **positividade** e a **dançabilidade**. O Gráfico 42 mostra o resultado do cálculo de **k** (número de clusters) e o motivo da escolha de quatro clusters.

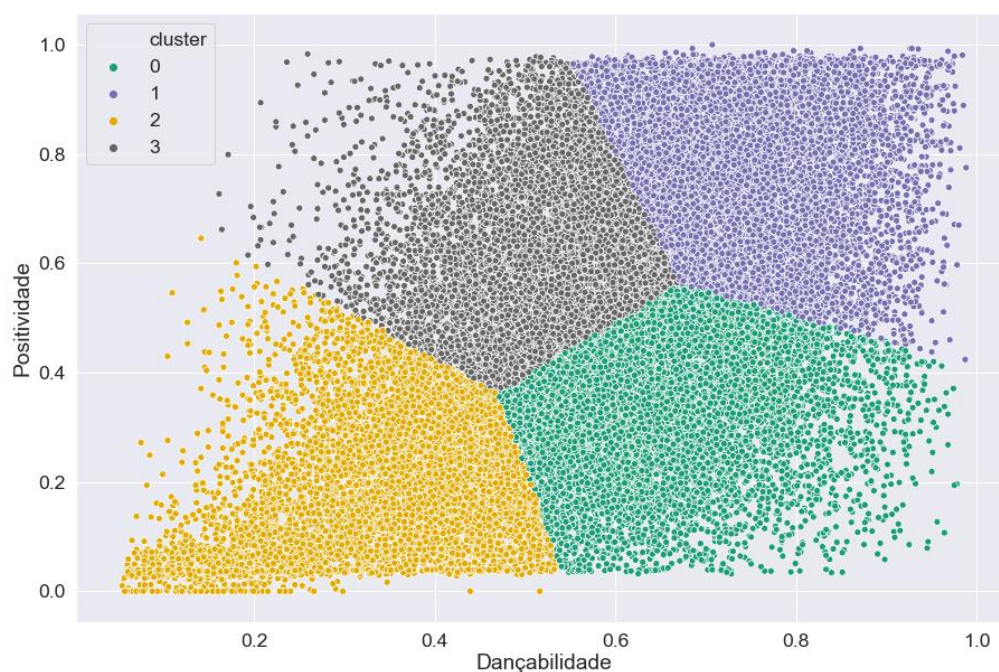
**Gráfico 42** – O Método *Elbow* (Cotovelo)



**Fonte:** Elaborado pelo autor, 2019.

A função do método retorna um gráfico, no qual conforme a curva diminui, podemos escolher um ponto do eixo x (k ou número de Clusters), onde a curva dos pontos diminui. E este ponto, no gráfico é k igual a 4.

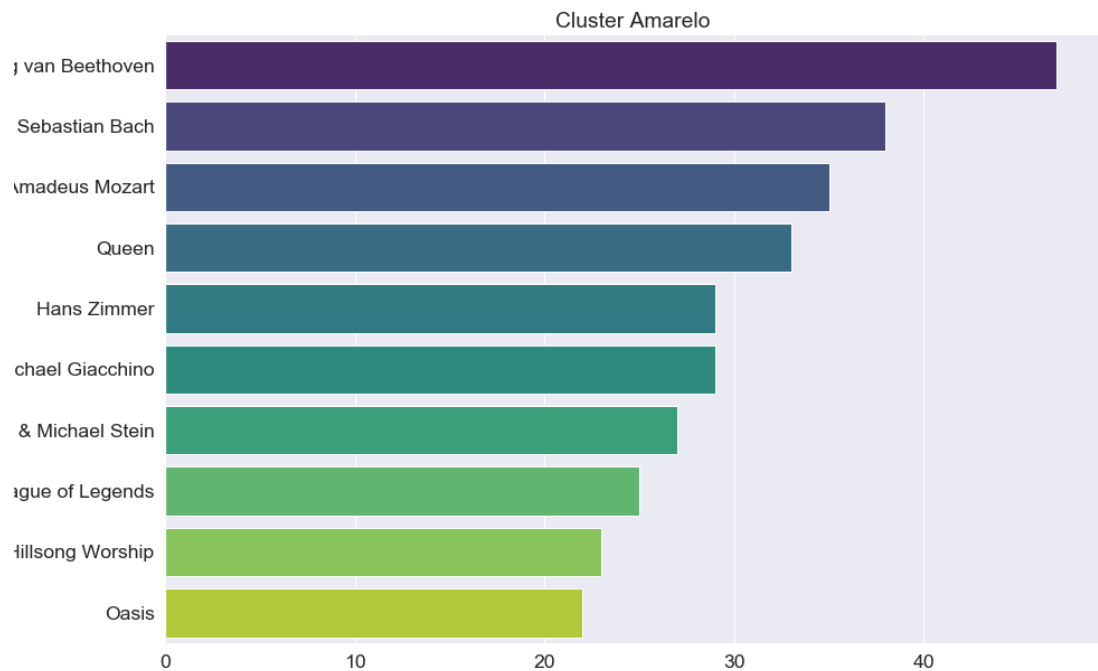
O Gráfico 43 mostra a clusterização para o k ou número de Clusters igual a 4.

**Gráfico 43 – Clusterização**

**Fonte:** Elaborado pelo autor, 2019.

Podemos observar, que na clusterização, os pontos maiores são relativos aos tons maiores, e os pontos menores, as tonalidades menores. Há uma quantidade maior de tonalidades maior. É uma característica dos dados analisados. Todos os países tem preferências por tonalidades maiores. Os grupos podemos considerá-los através da dançabilidade e positividade.

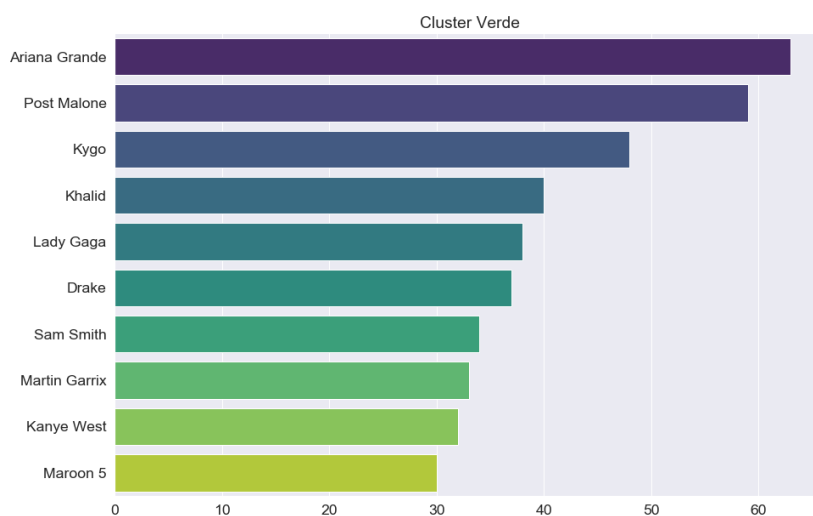
O cluster com a cor amarela, possui músicas que estão dispostas no nível de dançabilidade entre 0 a 0.5, sendo seu nível de positividade entre 0 e 0.5, também. Este grupo apresenta características de músicas menos dançantes. Podemos encontrar artistas como Beethoven, Bach, Mozart, Queen. Ou seja, artistas com estilos voltados à música clássica, ou com referência a ela. No Gráfico 44 podemos perceber essa predominância de artistas.

**Gráfico 44 – Artistas Cluster Amarelo**

**Fonte:** Elaborado pelo autor, 2019.

O agrupamento verde, possui um nível de dançabilidade entre 0.5 e 0.9, com uma positividade entre 0 e 0.5. O que caracteriza músicas dançantes, mas que, ao mesmo tempo, possam não ser tão positivas. Possui artistas como Ariana Grande, Post Malone, Lady Gaga. Artistas com estilos voltados ao pop e ao romantismo, como Sam Smith. Conforme Gráfico 45.

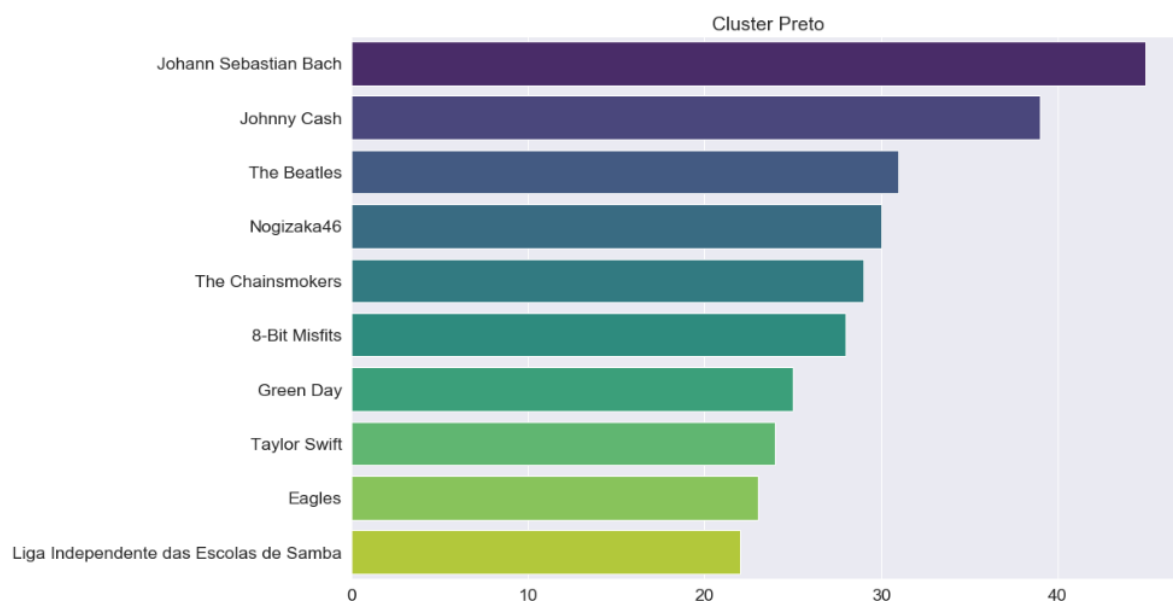
### Gráfico 45 – Artistas Cluster Verde



**Fonte:** Elaborado pelo autor, 2019.

Já o grupo preto, há um nível de dançabilidade que vai de 0.3 a 0.7. Apesar da dançabilidade baixa em alguns pontos, o conjunto de músicas deste cluster apresenta um nível de positividade acima de 0.5. Possui artistas conhecidos como Green Day, Bach, The Beatles. Há uma variedade de estilos neste clusters, pelo motivo da dançabilidade e positividade ser equilibrada. Podemos observar artistas do pop, samba, country, clássico. Conforme Gráfico 46.

### Gráfico 46 – Artistas Cluster Preto

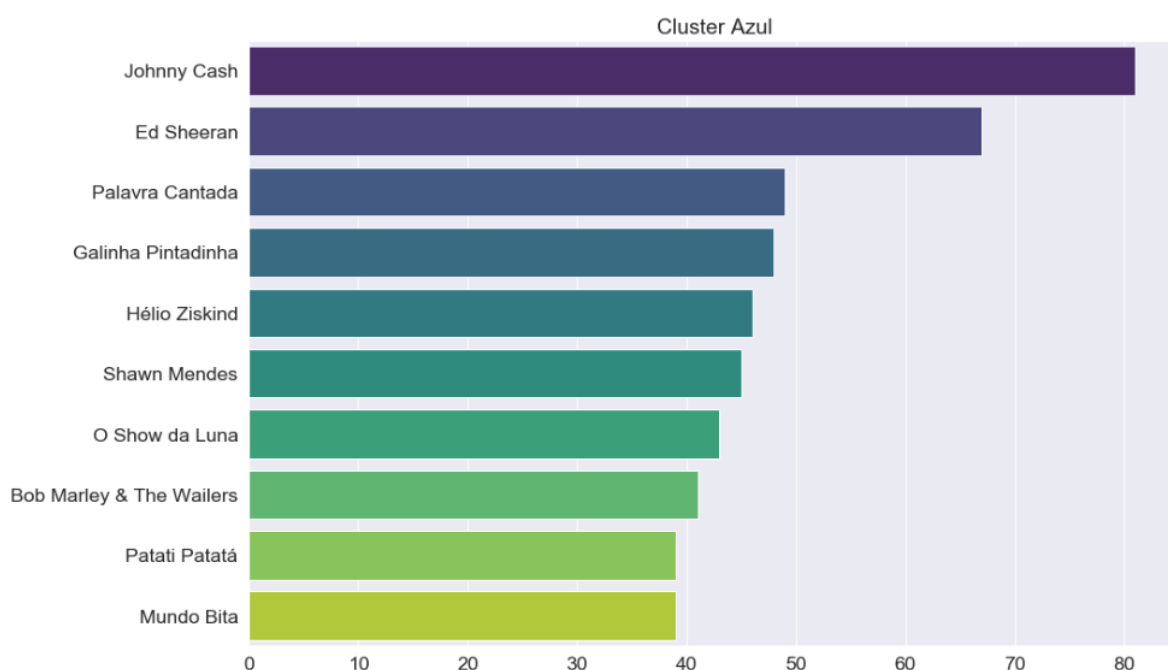


**Fonte:** Elaborado pelo autor, 2019.



O cluster azul é o que apresenta o maior nível de dançabilidade e positividade. Onde, nos dois atributos, os níveis iniciam a partir de 0.5. Tem como artistas Jhonny Cash, Turma da Mônica, O Show da Luna, Palavra Cantada. Possui músicas muito dançantes, o que traz consigo, músicas muito mais positivas. Há uma predominância de artistas infantis, acredito que pelo fato das características positivas. Mas além desse estilo infantil, possui estilos como o raggae, representado pelo Bob Marley & The Wallers. Podemos então afirmar, que as músicas mais dançantes, são voltados principalmente ao público infantil. Conforme Gráfico 47.

**Gráfico 47 – Artistas Cluster Azul**



**Fonte:** Elaborado pelo autor, 2019.

Portanto, diante do exposto, o objetivo desta seção foi analisar os dados provenientes do *Spotify* acerca das músicas contidas nas playlists dos países Brasil, França, Japão e Estados Unidos, de forma individual, mas ao longo da análise, também realizada de maneira global. Sendo assim, análise consistiu na identificação de alguns padrões de estilos e categorias musicais mais tocadas nesses países. Além de ter mostrado também os artistas mais ouvidos em cada país analisado. Puderam ser observado as características de três artistas: Johnny Cash, Ed Sheeran e Johann Sebastian Bach. Artistas consagrados em seus estilos musicais. Também, foi

mostrado um exemplo de clusterização. Dividindo os dados por padrões de dançabilidade e positividade, nos dando uma visão de mercado musical muito amplo.

## 5 CONSIDERAÇÕES FINAIS

Neste estudo, foi possível observar que os *streamings* de música, como o *Spotify*, constituem-se em um novo meio de consumir músicas de todos os estilos e categorias, bem como, agregam artistas diversificados de países distintos. Esse aspecto nos faz perceber o quanto a música passou por diversos processos para chegar ao ouvinte. Pois, antes desta evolução tecnológica presente hodiernamente, ouvia-se músicas por meio de instrumentos musicais, depois pelo rádio, LPs, Fita Cassete, CDs. Esses, de certo modo, tornaram-se obsoletos.

Com tantas mudanças, os *streamings* de música se tornaram uma ferramenta bastante significativa no modo de consumo musical, pois alguns usuários de plataforma, como o *Spotify*, podem ouvir música em qualquer lugar e momento sem estar diante de um aparato radiofônico, basta ter um celular com internet. O *Spotify* é uma empresa sueca que possui uma diversidade de músicas, as quais qualquer indivíduo pode acessar, criar playlists de acordo com seu gosto, além de interagir com outros usuários.

Analisando esses aspectos, percebeu-se a importância de estudar esse *streaming* para compreender quais as músicas, categorias, estilo e artistas são mais ouvidos na plataforma nos países Brasil, França, Japão e Estados Unidos. Sendo assim, buscou-se abordar, inicialmente, alguns conceitos sobre bancos de dados, MongoDB, Mineração de texto entre outros assuntos, com a finalidade de entender como funciona o processo de análise dos dados de músicas que foram obtidos.

Realizou-se a investigação através das playlists de cada país já mencionado, foi visto que há uma quantidade significativa de músicas nos países, por estilo, categoria e os artistas mais ouvidos. Observou-se que os Estados Unidos (US) contêm a maior concentração de músicas, seguido do Brasil.

Por conseguinte, observou-se, também, que cada país possui características específicas, desde os artistas ouvidos aos estilos, consideradas categorias pelo *Spotify*. As diferenças são pequenas entre cada um deles. No entanto, os artistas ouvidos são, de modo geral, distintos. Havendo artistas com maior fama internacional, como Ed Sheeran, que possui clássicos da música pop atual, Johann Sebastian Bach, na música clássica, Johnny Cash, no country, Lady Gaga, Taylor Swift, Queen.

O quão dançante uma música é, traz consigo energia, positividade. Que são atributos com alta correlação, ou seja, quanto maior o nível de **dançabilidade**, maior

será o nível de **positividade** e **energia** da música. O que pode caracterizar aspectos culturais do país. Para o Brasil, por exemplo, há uma predominância de artistas infantis, a exemplo, Galinha Pintadinha, Palavra Cantada, o que demonstra o pouco do que vemos nos dias atuais, as crianças tendo mais acesso à tecnologia, fazendo com que tenham, de certa forma, influenciado nas escolhas dos pais, principalmente mais jovens.

O nível de dançabilidade, de forma geral, possui um nível dançante, com o valor de 0.6. O que caracteriza um andamento moderado. O nível de energia também apresenta uma média de 0.6. E a positividade, está no nível de 0.5. É, portanto, um conjunto de músicas que podemos considerar dançante, positiva e com muita atividade, ou seja, energia. Podemos encontrar artistas dos mais variados, desde artistas pop, Ed Sheeran, há artistas clássicos, como Bach e Chopin. Os estilos, ou categorias, conforme é nomeado pelo Spotify, possui estilos como Sertanejo, Jazz, Pop, *Sleep*, a qual é caracterizada por músicas com andamentos lentos. Como exemplo a música Hush, da artista Luna Meguise, com um nível de dançabilidade 0.26, o que caracteriza, de fato, uma música com baixa dançabilidade, da categoria *Sleep*.

O nível de dançabilidade por país é equivalente, ou seja, possuem métricas aproximadas. Assim como os outros atributos. Dessa forma, podemos verificar que as diferenças, das músicas que estão no top de cada país, possuem pequenas diferenças, que ainda assim, podem caracterizar um país.

No Brasil, há a predominância de artistas infantis, como Palavra Cantada, Mundo Bitá, Turma da Mônica, entre outros. É um país conhecido atualmente pelo estilo Sertanejo Universitário, no entanto, nos dados analisados, há esta predominância por um estilo familiar.

Na França, há uma variedade de estilos bem maior que o Brasil, voltados ao clássico, hip-hop, jazz, rock, pop, entre outros. Podemos encontrar vários artistas, inclusive artistas brasileiros, Zezé Di Camargo & Luciano. Artistas brasileiros, que são ouvidos em outros países.

O Japão é um país que, assim como a França, passeia por vários estilos, em sua maioria estrangeira. Possui como estilo predominante o da cultura pop. Mas também estilos como o pop, rock, country. É o país onde a predominância de músicas clássicas foi maior, contendo artistas como Johann Sebastian Bach.

Nos Estados Unidos podemos perceber uma variedade de estilos, que vão desde o Jazz ao Hiphop, estilos que ganharam força através deste país. Podemos encontrar artistas como Ambjaay do hip-hop, ao Spyro Gyra que é voltado para o Jazz.

Logo, depreende-se que a facilidade de acesso às músicas no *Spotify* torna possível uma gama de análises, das quais podem influenciar o direcionamento de propagandas, identificação de grupos específicos de ouvintes, entre outros aspectos. Dessa forma, pode ser aplicado técnicas de Aprendizagem de Máquina para identificar, a partir das características de uma música, em qual localidade poderia obter mais sucesso.

## REFERÊNCIAS

AMARAL, Fernando. **Introdução à Ciência de dados: Mineração de dados e Big Data**. Rio de Janeiro: Alta Books, 2016, 320p.: il.

BACH, Johann Sebastian. **Gênio traduz mistérios do sagrado**. Coleção Folha de Músicas Clássicas. Biografia. Folha Online. Disponível em: <https://musicaclassica.folha.com.br/cds/17/biografia.html> Acesso em: 25 de nov. 2019.

BENITO, Sandy et al. **Music Data Mining using audio Features Extracted from Spotify**. St. Thomas University. School of Science. Miami Gardens, FL. Disponível em [https://www.reisanar.com/files/poster\\_Benito\\_Carr\\_v5.pdf](https://www.reisanar.com/files/poster_Benito_Carr_v5.pdf) Acesso em: 15 de out. 2019.

BIOGRAFIA. **Ed Sheeran**. Vagalume. Disponível em: <https://www.vagalume.com.br/ed-sheeran/biografia/> Acesso em: 15 de out. 2019.

BRAGA, Vitor et al. **O streaming de música como um estímulo para a ampliação do consumo musical: um estudo do Spotify**. In: XXXIX CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO. 05 a 09 de setembro de 2016. Intercom – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação. São Paulo, SP, 2016.

CARVALHO FILHO, José Adail. **Mineração de textos: análise de sentimentos utilizando Tweets referentes à Copa do Mundo**. 2014. 46f.: il. Monografia (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Engenharia de Software, Quixadá, 2014.

COELHO, Taysa. **Fita cassete completa 50 anos; veja o que mudou em cinco décadas**. 2013. Disponível em: <https://www.techtudo.com.br/noticias/noticia/2013/09/fita-cassete-completa-50-anos-veja-o-que-mudou-em-cinco-decadas.html> Acesso em 24 de nov. 2019.

ELMASRI, Ramez; NAVATHE, Shamkant B. **Sistemas de banco de dados**. Tradução: Daniel Vieira; revisão técnica Enzo Scraphim e Thatyana de Faria Piola Scraphim. 6. ed. - São Paulo: Pearson Addison Wesley, 2011.

EVERITT, Brian S. et al. **Cluster Analysis**. 5<sup>th</sup> Edition. John Wiley & Sons, Ltd. 2011. Disponível em: <https://books.google.com.br/books?id=w3bE1kqd-48C&printsec=frontcover&hl=pt-BR#v=onepage&q&f=false>. Acesso em: 15 de out. 2019.

FÉTIS, François-Joseph. *Traité complet de la théorie et de la pratique de l'harmonie: contenant la doctrine de la science et de l'art*. 8. ed. Paris: G. Brandus et S. Dufour, 1864. In: FREITAS, Sérgio Paulo Ribeiro de; LIBRELOTTO, Edilamar Ilha. **O conceito de tonalidade segundo François-Joseph Fétis**: tradução comentada de um artigo de divulgação de Robert Wangermée. DAPesquisa, v.11, n.16, p254-275, agosto 2016.

GAN, Guojun et al. **Data Clustering: Theory, Algorithms, and Applications**. ASA-SIAM Series on Statistics and Applied Probability. SIAM, Philadelphia, ASA, Alexandria, VA, 2007.

GRUS, Joel. **Data Science do Zero**. Rio de Janeiro: Alta Books, 2016. Disponível em:  
file:///C:/Users/maris/Desktop/ADRIANO/capitulo\_de\_amostra\_Data\_Science\_do\_Zero.pdf. Acesso em: 26 out. 2019.

KISCHINHEVSKY, Marcelo et al. **Música infinita: serviços de streaming como espaços híbridos de comunicação e consumo musical**. Compós. Associação Nacional dos Programas de Pós-Graduação em Comunicação. Disponível em: [http://www.compos.org.br/biblioteca/test\\_2860.pdf](http://www.compos.org.br/biblioteca/test_2860.pdf). Acesso em: 26 out. 2019.

LAZZARIN, Lilian N. A. **Técnica para mineração de textos na análise de sentimentos: Um estudo de caso em uma Instituição de Ensino Técnico, Tecnológico e Superior**. 2017. Monografia (II Curso de Especialização em Banco de Dados) – Universidade Tecnológica Federal do Paraná. Pato Branco, 2017.

LÓSCIO, Bernadette Farias et al. **NoSQL no desenvolvimento de aplicações Web colaborativas**. 2011. Disponível em:  
[https://www.addlabs.uff.br/sbsc\\_site/SBSC2011\\_NoSQL.pdf](https://www.addlabs.uff.br/sbsc_site/SBSC2011_NoSQL.pdf) Acesso em: 26 set. 2019.

MONDELLI, Maria Luiza Botelho et al. O que os países escutam: Analisando a Rede de Gêneros Musicais ao Redor do Mundo. **Anais do VII Brazilian Workshop on Social Network Analysis and Mining – BraSNAM**, 2018. Disponível em:  
<https://www.lncc.br/~lgadelha/publications/brasnam-2018.pdf> Acesso em 24 de nov. 2019.

PRIOLLI, Maria Luisa de Mattos. **Princípios básicos da música para a juventude**. 48º ed. Revista e Atualizada. Rio de Janeiro: Casa Oliveira de Música – LTDA, 2006.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. – Novo Hamburgo, RS: Feevale, 2013. Livro eletrônico. Disponível em:  
[https://www.academia.edu/36705797/PRODANOV\\_Metodologia\\_do\\_Trabalho\\_Cientifico?auto=download](https://www.academia.edu/36705797/PRODANOV_Metodologia_do_Trabalho_Cientifico?auto=download) Acesso em 29 de out. 2019.

RIBEIRO, Eduardo. **Qual é o papel do CD no mercado da música hoje?** 2017. Disponível em: [https://www.vice.com/pt\\_br/article/wj8ygn/cd-mercado-fonografico-estudando-cena-3](https://www.vice.com/pt_br/article/wj8ygn/cd-mercado-fonografico-estudando-cena-3) Acesso em 24 de nov. 2019

RODRIGUES, Aíquis et al. **Spotify em Foco: Um Estudo de Caso sobre Sistemas para a Terceira Plataforma Computacional**. 2016. Disponível em:  
<http://labbd.ufrj.br/ersi2016/wp-content/uploads/2016/07/Spotify-em-Foco-Um-Estudo-de-Caso-sobre-Sistemas-para-a-Terceira-Plataforma-Computacional.pdf> Acesso em 29 de out. 2019

SANTOS, Mylena C. Gadelha; RAMOS, Rebecca Costa. **Aplicativos de música: o Spotify, as mudanças no mercado fonográfico e os filtros-bolha**. In: XXXIX CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO. 05 a 09 de setembro de 2016. Intercom – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação. São Paulo, SP, 2016. Disponível em: <http://portalintercom.org.br/anais/nacional2016/resumos/R11-2913-1.pdf> Acesso em: 26 out. 2019.

SANTOS, Hadrizia. **Análise das músicas do Spotify**. Publicado em: 27 de maio de 2018. Disponível em: [https://rstudio-pubs-static.s3.amazonaws.com/392416\\_a7386ee89fdf41039f3abe7deb4929e1.html](https://rstudio-pubs-static.s3.amazonaws.com/392416_a7386ee89fdf41039f3abe7deb4929e1.html) Acesso em: 25 de nov. 2019.

SOLAGNA, Emanuele Araujo; LAZZARETTI, Alexandre Tagliari. **Um estudo comparativo entre O mongodb e o postgresql**. 2016. 23f. Trabalho de Conclusão de Curso (Graduação em Tecnólogo em Sistemas para Internet). Curso de Tecnologia em Sistemas para Internet. Instituto Federal Sul-Rio-Grandense, Campus Passo Fundo, 2016. Disponível em: <https://painel.passofundo.ifsul.edu.br/uploads/arq/201607111805501015914198.pdf> Acesso em: 26 out. 2019.

TECMUNDO. **A evolução do armazenamento de músicas [infográfico]**. 2012. Disponível em: <https://www.tecmundo.com.br/infografico/30658-a-evolucao-do-armazenamento-de-musicas-infografico-.htm> Acesso em 24 de nov. 2019.

WAZLAWICK, Raul Sidnei. Uma Reflexão sobre a Pesquisa em Ciência da Computação à Luz da Classificação das Ciências e do Método Científico. **Revista de Sistemas de Informação da FSMA**, n. 6, 2010, p. 3-10.

WIKIPÉDIA. **Johnny Cash**. Biografia. Disponível em: [https://pt.wikipedia.org/wiki/Johnny\\_Cash#Biografia](https://pt.wikipedia.org/wiki/Johnny_Cash#Biografia) Acesso em 24 de nov. 2019.

WUNDERVALD, Bruna; TRECENTI, Julio. Análise de dados musicais no R. **IV Seminário Internacional de Estatística com R**. Maio, 2019. Disponível em: <http://brunaw.com/shortcourses/IXSER/pt-br/pres-pt-br.html#2> Acesso em: 25 de nov. 2019.



