FACULDADE DE TECNOLOGIA DE SÃO JOSÉ DOS CAMPOS

FATEC PROFESSOR Jessen Vidal

VINÍCIUS PIETRO CARMO DOS SANTOS

ANÁLISES DE PLAYLISTS DO SPOTIFY

Orientador: Fernando Masanori Ashikaga

São José dos Campos

2020

SUMÁRIO

[1 Introdução 3](#_Toc44848594)

[1.1 Definição do problema 3](#_Toc44848595)

[1.2 Objetivo 3](#_Toc44848596)

[2 Desenvolvimento 3](#_Toc44848597)

[2.1 Detalhes do Desenvolvimento 3](#_Toc44848598)

[2.1.1 Autenticação do Token de Usuário 4](#_Toc44848599)

[2.1.2 Obtenção do Conteúdo da Playlist 4](#_Toc44848600)

[2.1.3 Obtenção dos “audio\_features” 5](#_Toc44848601)

[2.1.4 Exibição dos Dados Obtidos 6](#_Toc44848602)

[2.2 Casos de Análise 8](#_Toc44848603)

[2.2.1 As minhas músicas do Spotify, são entediantes? 8](#_Toc44848604)

[2.2.2 Comparação das Playlist Top 50 Brasil e Top 50 USA 9](#_Toc44848605)

[2.2.3 Comparação da Playlist Top 50 Brasil, pré e durante pandemia COVID-19 12](#_Toc44848606)

[2.2.4 Análise da Playlist Sertanejo Sofrência 13](#_Toc44848607)

[2.2.5 Evolução Musical da banda Evanescence 14](#_Toc44848608)

[3 Conclusão 16](#_Toc44848609)

[4 Referências 16](#_Toc44848610)

# Introdução

Este Capítulo apresenta o problema a ser resolvido e o objetivo deste trabalho.

## Definição do problema

Não é de hoje que se ouve os termos Mineração de Dados e Análise de Dados, ambos processos de identificação de padrões e tendências principalmente utilizados em empresas que geram e consomem uma grande carga de dados, gerados principalmente pelos usuários que acessam e utilizam seus aplicativos, gerando assim tendências. Ao pesquisar mais sobre a análise de dados, me deparei com um artigo sobre como as músicas da dupla Sandy e Junior, ficaram mais tristes com o tempo[[1]](#endnote-1). Lendo o artigo me perguntei se era possível aplicar uma análise parecida, porém analisando outros casos como: Seria possível comparar duas playlists? Ou até mesmo, seria possível realizar uma análise da evolução musical de um artista ou grupo.

## Objetivo

O objetivo deste trabalho consiste em estudar a aplicação de algoritmos de mineração de dados em playlists do Spotify, visando analisar os dados obtidos para conseguir identificar gostos musicais e tendências dentre outros casos de análise, a partir de playlist públicas e privadas do Spotify.

# Desenvolvimento

## Detalhes do Desenvolvimento

O código foi desenvolvido em Python, uma linguagem de programação simples e com sintaxes de fácil uso, utilizada preferencialmente em aplicações de desenvolvimento rápido. Já a plataforma de desenvolvimento foi o Jupyter Notebook, uma aplicação web que permite a visualização e edição de códigos, principalmente os que envolvem mineração e visualização de dados, transformação de dados e muito mais.

De um modo geral o código pode ser divido em quatro partes, Autenticação do Token de Usuário, Obtenção do conteúdo da playlist, Obtenção dos “audio\_features” e Exibição dos dados obtidos.

### Autenticação do Token de Usuário

Para realizar a autenticação do Token de Usuário, primeiramente deve-se configurar o seu usuário no Spotify for Developers, para obter os valores de *client\_id* e *client\_secret.* Notando que o valor *username* equivale ao código do usuário no aplicativo do Spotify. O código descrito na Figura 1, é responsável por realizar esta autenticação, para que os métodos descritos nas Figuras 2 e 3, possam funcionar corretamente.

A autenticação é realizada, conforme demonstrado na Figura1, nas linhas 20 a 25, onde caso as credenciais de acesso (*client\_id* e *client\_secret*) equivalerem a um cadastro existente no site Spotify for Developers, um prompt será exibido, indicando que o Token de acesso foi autorizado para aquelas credenciais, permitindo o uso de funções onde a autenticação é necessária.

**Figura 1 –** Autenticação token de usuário

**

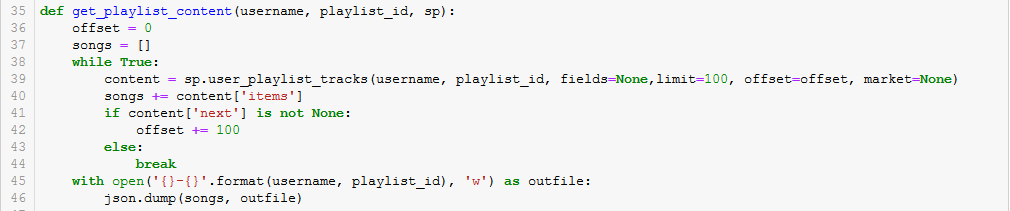
### Obtenção do Conteúdo da Playlist

O conteúdo de uma playlist, equivale a músicas/faixas que a compõe.

O método responsável por obter todo o conteúdo da playlist e salvá-los em um JSON, pode ser visualizado na Figura 2. De um modo geral o método pode ser explicado da seguinte forma:

* Na linha 35, temos a definição dos métodos e quais os parâmetros que irá receber, neste caso, o método recebe os parâmetros *username, playlist\_id* e *sp;*
  + Os parâmetros *username* e *sp*, servirão para revalidar o token durante a execução do método;
  + Já o parâmetro *playlist\_id*, indica para o código, qual a playlist que deve ter seus dados obtidos;
* Nas linhas 38 a 43, o código força um loop para obter o conteúdo da playlist em questão;
* Na linha 44, quando não houver mais conteúdo para obter, o código interrompe o loop;
* E na linha 45, salva os dados obtidos em um JSON.

**Figura 2 –** Método de obtenção dos conteúdos da playlist



### Obtenção dos “audio\_features”

O método apresentado na Figura 3, demonstra a obtenção dos dados, chamados de audio\_features, a partir do método apresentado na Figura 2.

Como demonstrado na Figura 3, na linha 57 o código força um loop para obter os dados da playlist em questão, parecido com o loop mencionado na seção 2.1.2. Após sair desse loop, quando não houver mais conteúdo na playlist, na linha 64 o código entra em um outro loop, dessa vez passando por cada música/faixa da playlist, obtendo em cada iteração do loop, as *audio\_features*. Obtendo as *audio\_features* da faixa, na linha 81 os dados obtidos são formatados para o formato Data Frame, de modo que fique mais fácil ler estes dados futuramente. Na linha 87 é salvo os dados, formatados em Data Frame, em uma planilha do Excel, após salvar o código, ainda dentro do loop, irá realizar os mesmos passos para todas as músicas/faixas subsequentes da playlist, até que não tenha mais nenhuma música.

**Figura 3 –** Método de obtenção das audio\_features de cada música da playlist



### Exibição dos Dados Obtidos

Para exibir os dados obtidos, foi utilizada a biblioteca Pandas, disponível para Python. Os dados obtidos[[2]](#endnote-2), demonstrados na Figura 4, podem ser compreendidos da seguinte forma:

Energy: medida de 0 a 1 que representa a percepção de intensidade e atividade musical. Faixas com o valor mais próximo de 1 são consideradas mais frenéticas enquanto faixas com o valor mais próximo de 0 são consideradas mais calmas;

Liveness: detecta a presença de uma audiência no áudio. Valores mais próximos de 1 indicam que a faixa foi gravada ao vivo;

Tempo: representa quantas batidas por minuto a faixa possuí;

Speechiness: detecta a presença de palavras faladas em uma faixa. Valores acima de 0,6 representam faixas que consistem inteiramente de palavras faladas (podcasts), valores entre 0,6 e 0,3 indicam faixas que contém músicas e palavras faladas (raps) e valores abaixo de 0,3 representam faixas musicais;

Acousticness: uma medida de 0 a 1, que indica se a faixa é acústica ou não. 0 representando não e 1 representando sim;

Instrumentalness: medida que identifica a presença de vocais em uma faixa;

Time\_Signature: uma medida que verifica a visão geral da assinatura de tempo de uma faixa;

Danceability: medida de 0 a 1, indicando se a faixa é dançável ou não. 0 representando não e 1 representando sim;

Key: representa a nota musical mais presente na música. A representação é baseada na notação de números inteiros (Pitch Class Notation);

Duration\_ms: representa a duração em milissegundos de uma faixa;

Loudness: a medida de altura relativa de uma faixa, medida em decibéis;

Valence: medida de 0 a 1 descrevendo a positividade musical transmitida pela faixa. Valores mais próximos de 1 transmite emoções positivas enquanto faixas com valor mais próximos de 0 transmite emoções negativas;

Mode: indica a modalidade da música sendo representados como Maior ou Menor;

Type: o tipo do objeto, neste código, sempre “audio\_features”;

URI: o código URI da faixa no Spotify.

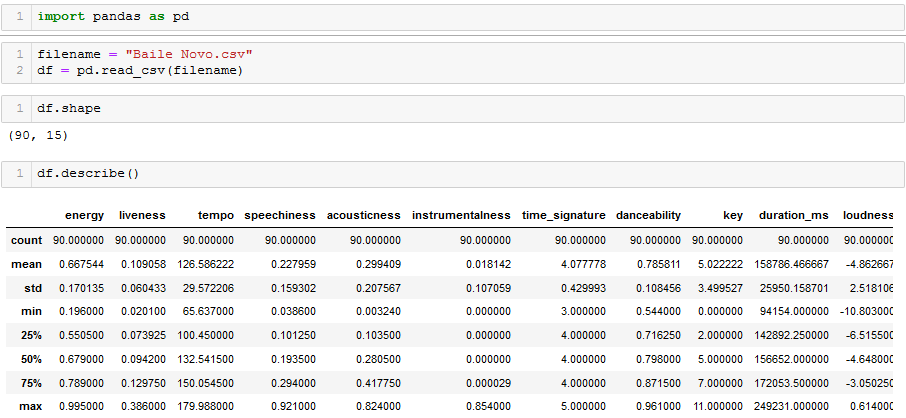
A exibição é feita utilizando as funções de exibição Data Frames, fornecidas pelo Pandas, mostrando em uma tabela os seguintes valores, onde os utilizados para a análise serão:

Min: o menor valor da audio\_feature presente na playlist;

Mean: valor médio da audio\_features em questão;

Max: o maior valor da audio\_feature presente na playlist.

**Figura 4 –** Exemplo de exibição dos dados obtidos



## Casos de Análise

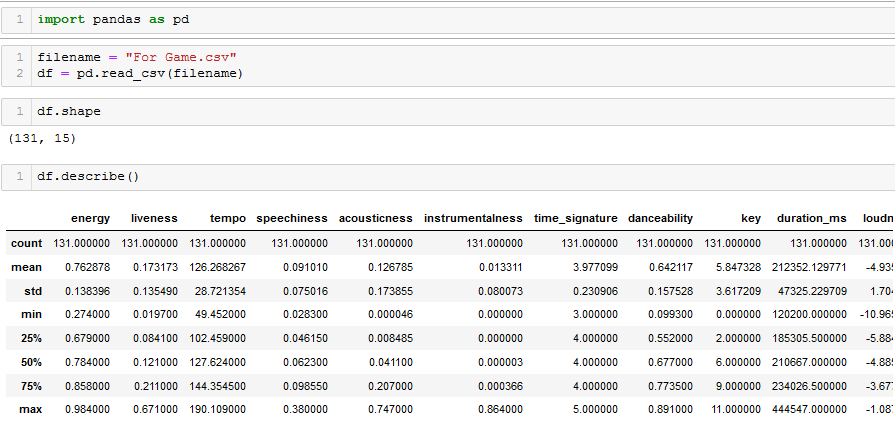
### As minhas músicas do Spotify, são entediantes?

Sempre ouvi de uma amiga que a minha playlist era muito "chata" e "entediante", por conter músicas como *Prince of Peace*, *Dangerous Woman, Don’t Think Twice* e *Phoenix*. Nessa análise vamos ver se essa percepção sobre a minha playlist é verdadeira ou não. Analisando de forma geral a minha playlist, pelos resultados demonstrados na Figura 5 a playlist For Game, é uma playlist bem intensa, bem dançável com um humor geral bem no meio termo, nem tão feliz nem tão triste, com pouca presença acústica entre as faixas. Já a playlist Zen, considerada por essa amiga como animada, pela presença de músicas como *O Sol, Meu Abrigo* e *Me* Namora, é uma playlist mais calma porém também bem dançável, novamente nem tão feliz e nem tão triste, já a presença acústica nas faixas é bem mais visível, fatos evidenciados pelos dados exibidos na Figura 6.

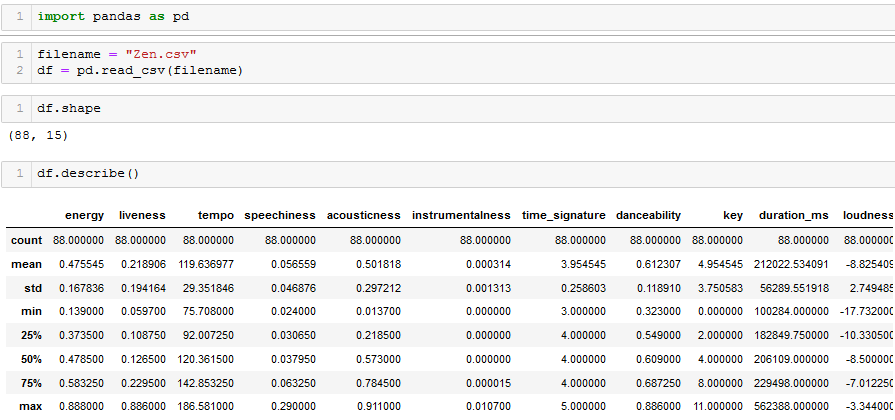
Comparando ambas as playlists, como demonstrado na figura 7, podemos ver a disparidade entre os campos, comprovando a diferença de intensidade e presença de acústico das duas playlists.

De um modo geral, minha preferência por música pode ser resumida em faixas bem intensas, dançáveis e com pouca presença de tons e sons acústicos, enquanto a preferência da minha amiga, são músicas calmas, porém dançáveis e com uma presença média de tons e sons acústicos. Indicando que por minha playlist ser mais intensa e gosto pessoal dessa minha amiga ser por músicas mais calmas, essa sensação de tédio, surge justamente pela minha playlist ser “over” demais para o gosto dela.

**Figura 5 –** Dados obtidos da playlist For Game



**Figura 6 –** Dados obtidos da playlist Zen

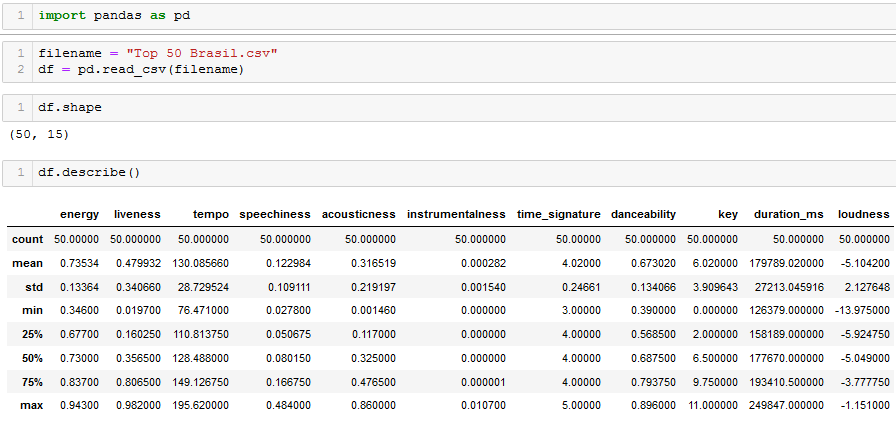


**Figura 7 –** Exibição gráfica dos dados obtidos pela comparação entre as playlists For Game e Zen

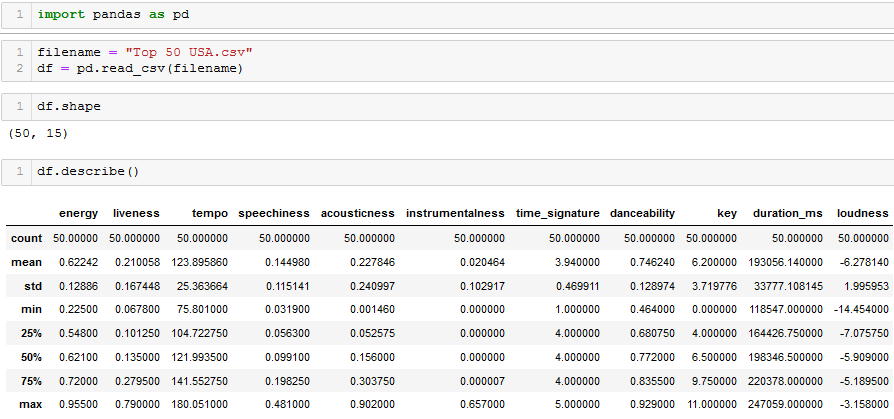
### Comparação das Playlist Top 50 Brasil e Top 50 USA

Analisando as Top 50 músicas mais ouvidas no Brasil e Estados Unidos, conseguimos ver algumas diferenças significativas nos dados obtidos dessas playlists. Primeiramente podemos observar que a playlist brasileira, como demonstrado na Figura 8, contém faixas mais intensas, parte bem importante da cultura e assinatura musical brasileira, fato que é evidenciado pelos gêneros mais comuns ouvidos no país (Funk, Eletrônica, Samba, entre outros). Também conseguimos perceber a presença de uma tendência que vem surgindo nas músicas brasileiras, que são as gravações ao vivo, utilizada principalmente por cantores, duplas e bandas Sertanejas[[3]](#endnote-3). Já na playlist Americana, demonstrado na Figura 9, temos a presença levemente intensas, fato evidenciado pelo crescente aumento das reproduções dos gêneros Hip-Hop R&B. Além disso notamos uma grande diferença na presença de faixas ao-vivo, tendência que diferentemente da cultura no Brasil não é tão presente nos Estados Unidos. De um modo geral, a tendência brasileira quando comparada com a americana como evidenciado na Figura 10, é mais intensa e com mais elementos ao-vivo em suas faixas, enquanto a americana, é um pouco suavizada e mais dançáveis.

**Figura 8 - –** Dados obtidos da playlist Top 50 Brasil (Abril - 2020)



**Figura 9 -** Dados obtidos da playlist Top 50 USA (Abril - 2020)



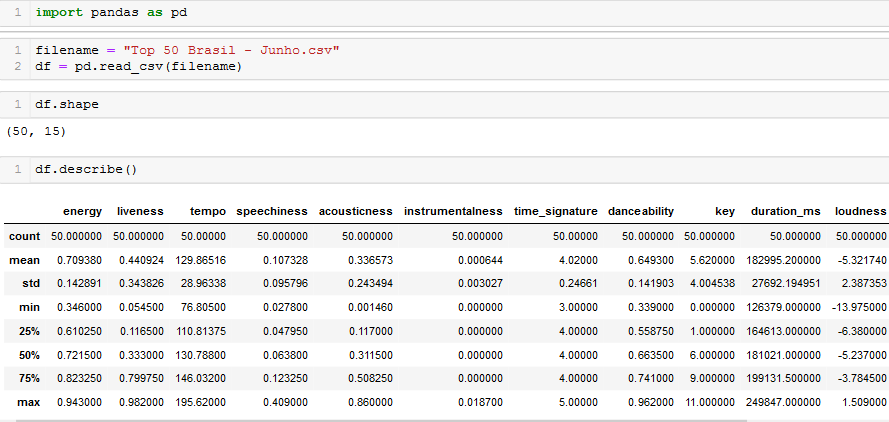
**Figura 10 -** Exibição gráfica dos dados obtidos pela comparação entre as playlists Top 50 Brasil e Top 50 USA

### Comparação da Playlist Top 50 Brasil, pré e durante pandemia COVID-19

Diversos costumes, rotinas e tendências foram afetadas direta e indiretamente pela pandemia da COVID-19. A música foi um desses costumes afetados. Forçando artistas a procurarem novos meios de gravação, é comum que tenha havido uma alteração nas tendências atuais, não só no Brasil, como no Mundo.

Porém ao analisarmos e compararmos os dados pré e durante pandemia, como demonstrado nas Figuras 8 e 11 respectivamente, conseguimos observar que não houve uma alteração tão significativa da tendência brasileira, mantendo sua característica base, músicas intensas, dançáveis e com a presença de faixas ao-vivo, fato evidenciado pela comparação entre as playlists, como visto na Figura 12. Isso se dá pelo fato de artistas brasileiros estarem lançando de poucas a quase nenhuma música nesta época, optando por fazer transmissões ao vivo de músicas já gravadas em outras plataformas, como Youtube[[4]](#endnote-4).

**Figura 11 -** Dados obtidos da playlist Top 50 Brasil em Junho

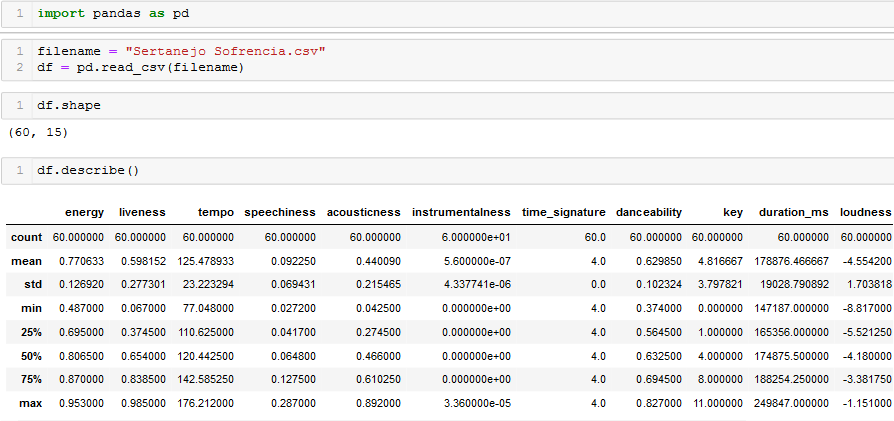


**Figura 12 -** Exibição gráfica dos dados obtidos pela comparação entre as playlists Top 50 Brasil em Abril e Top 50 Brasil em Junho

### Análise da Playlist Sertanejo Sofrência

Popularmente conhecido e chamado de Sertanejo Sofrência, este sub gênero pelo próprio nome, demonstra uma pré-disposição do sentimento passado pelas suas faixas, que seria o sofrimento[[5]](#endnote-5). Mas será que sofrimento é exatamente a sensação passada por essas músicas? Observando de um primeiro momento os dados obtidos, facilmente podemos responder que apesar de não serem as músicas que mais passam uma sensação de felicidade, estão longe de passar também uma sensação de sofrimento, como evidenciado pelo valor de valência (62). Porém essa é a visão que nos é passada pela playlist, então mesmo que algumas dessas faixas possuam a valência baixa, a playlist em si tem um valor razoavelmente na média.

**Figura 13** - Dados obtidos da playlist Sertanejo Sofrência



**Figura 14 -** Exibição em gráfico das audio\_features obtidas da playlist Sertanejo Sofrência

### Evolução Musical da banda Evanescence

Utilizando os dados obtidos dos álbuns: Fallen, The Open Door, Evanescence, Lost Whispers, Synthesis e Synthesis Live, todos da banda Evanescence, podemos obter algumas análises bem interessantes, sobre a trajetória e evolução musical.

**Figura 15 -** Exibição em Gráficos das audio\_features dos álbuns do grupo Evanescence

O grupo musical Evanescence, surgiu no mundo da música com o albúm Fallen trazendo músicas como *Going Under, Bring Me to Life e My Immortal.* De um modo geral a banda no começo possuía um estilo de música bem único, a voz da cantora Amy Lee sempre atrelada a presença de instrumentos acústicos, como piano, dando uma impressão mais sombria e obscura, como pode-se observar nos índices de energia e valência (69 e 25, respectivamente) na Figura 15. O albúm The Open Door, continua com a mesma premissa, músicas pesadas, cheias de energia, porém ainda não chegam a ser dançantes, mostrando que energia e dançabilidade não estão necessariamente atreladas. Tendo como maiores sucessos deste albúm temos *Lithium, Call me When you're Sober e Lacrymosa*, podemos observar nessas faixas o conceito utilizado, batidas mais fortes e maior presença da voz da cantora.

O álbum Evanescence recebe o próprio nome da banda, por se tratar de um albúm que significou uma reinvenção de como seriam as músicas escritas e apresentadas pelo grupo, deste álbum para o futuro. Ainda apostando em músicas pesadas e sombrias, como demonstrado na Figura 15, com batidas fortes, porém com uma diferença bem significativa quando comparado com seus antecessores, o álbum apresenta a menor presença de música acústica. Enquanto possuí o maior índice de valência (28), o álbum não necessariamente é "feliz". Com a abertura para um novo conceito com o álbum Evanescence, o grupo segue o mesmo conceito, porém ainda realizando algumas alterações em seu estilo musical, o que fica evidente em Lost Whispers. Atingindo até o momento o menor índice de energia (60) e valência (23), nos indicando que álbum se apresenta como mais sombrio e mais calmo, apostando um pouco em uma apresentação mais fixa de instrumentalismo dentro de suas faixas, onde antes em outros álbuns o refrão era apresentado única e especificamente pela voz da cantora, neste álbum em faixas como *Together Again, Secret Door e A New Way to Bleed*, temos a apresentação de refrãos quase que compostos por apenas instrumentos, reforçado pelo álbum registrar o maior índice instrumental (4) entres os álbuns antecessores.

Já o álbum Synthesis e sua versão Ao Vivo, mostram novamente uma virada no estilo musical do grupo. Respeitando a premissa de "reinventar" suas músicas mais famosas e aclamadas pelo público, como *Bring Me to Life, Lithium e My Immortal* a banda utiliza menos a presença da voz potente da cantora, como demonstrando na Figura 15 e evidenciado pelo maiores índices acústico e instrumentais (70 e 15, respectivamente), optando por utilizar um estilo de música mais acústico, porém nunca abandonando a voz dramática de Amy Lee. Em termos técnicos o álbum Synthesis de modo geral, apresenta músicas mais leves, aumentando a presença, em alguns trechos de suas músicas, de instrumentos como piano e guitarra. Também distanciando o reconhecimento de um rock dramático para um rock mais leve. As diferenças mais gritantes observadas entre Synthesis e seu gêmeo Synthesis Live, na Figura 15, como os índices acústico (70 e 50) e instrumental (15 e 0,1), são explicados pela alteração no tipo de apresentação, quando gravadas em estúdio, as faixas de Synthesis tem foco em toda a obra, já enquanto na apresentação ao vivo, o centro das atenções é a capacidade vocal de Amy Lee.

# Conclusão

Diante dos casos apresentados, pode-se concluir que com Mineração e Análise de Dados de playlists do Spotify, é possível realizar diversos tipos de análise, como análises acadêmicas apresentadas neste trabalho, ou análises de mercado que considerariam outros dados como duração das músicas e nota musical mais presente nas faixas.

A execução bem sucedida deste trabalho, se deve a duas bibliotecas da linguagem Python, a Pandas que permitiu a utilização de estrutura em Data Frame para salvar e visualizar os dados obtido, e a biblioteca Spotipy que permitiu a obtenção dos dados do aplicativo.

# Referências

1. https://www1.folha.uol.com.br/ilustrada/2019/08/musicas-de-sandy-junior-ficaram-mais-tristes-com-o-tempo-mostra-pesquisa.shtml, acessado em 29/06/2020 às 16:32:00 [↑](#endnote-ref-1)
2. https://developer.spotify.com/documentation/web-api/reference/tracks/get-several-audio-features/, acessado em 30/06/2020 às 19:32:00 [↑](#endnote-ref-2)
3. https://www1.folha.uol.com.br/ilustrada/2019/11/brasil-e-o-pais-que-mais-escuta-musica-ao-vivo-no-mundo-todo.shtml, acessado em 26/06/2020 às 16:32:00 [↑](#endnote-ref-3)
4. https://exame.com/revista-exame/o-mundo-e-uma-live, acessado em 26/06/2020 às 17:12:00 [↑](#endnote-ref-4)
5. https://www1.folha.uol.com.br/ilustrada/2019/06/pe-na-bunda-e-o-que-mais-faz-brasileiros-conhecerem-novas-musicas.shtml acessado em 28/06/2020 às 10:43:00 [↑](#endnote-ref-5)