**PROJETO CONCLUSÃO DA DISCIPLINA**

Localização e Utilização da Informação

Professores: Felipe Borges & Evelyn Batista

Interface gráfica do usuário

Descrição gerada automaticamente

***Cristóvão Augusto Pessanha de Souza Júnior***

***Flávia Koreeda***

***BI MASTER***

Sumário

[1 INTRODUÇÃO 3](#_Toc80637308)

[1.1 API 3](#_Toc80637309)

[1.2 WEB SCRAPPING 3](#_Toc80637310)

[2 OBJETIVO 4](#_Toc80637311)

[2.1 - Extrair os dados (API ou Scraping). 4](#_Toc80637312)

[2.2 - Alocar essas informações em um DataFrame. 4](#_Toc80637313)

[2.3 - Executar a operação de ordenação no DataFrame em um atributo à sua escolha. 4](#_Toc80637314)

[2.4 - Executar a operação de filtragem no DataFrame à sua escolha. 4](#_Toc80637315)

[3 DESENVOLVIMENTO 5](#_Toc80637316)

[3.1 – Carregar Bibliotecas e Cadastramento da WEB API e obtenção do ID e KEY. 5](#_Toc80637317)

[3.2 – Definir artista e realizar a raspagem com a tecnologia API. 5](#_Toc80637318)

[3.3 – Raspar as URI’s de todos os álbuns dentro do Spotify. 6](#_Toc80637319)

[3.4 – Selecionar todas as músicas de cada álbum. 7](#_Toc80637320)

[3.5 – Adicionar cada música ao dicionário do álbum correspondente. 7](#_Toc80637321)

[3.6 – Extrair e exibir as características de cada faixa. 7](#_Toc80637322)

[3.6 – Recursos temporal organizador para tornar a operação de extração das características de cada faixa mais dinâmica e rápida. 9](#_Toc80637323)

[3.7 – Criando o DataFrame para o artista, álbum, faixa, features e popularidade. 9](#_Toc80637324)

[3.8 – Eliminar as músicas duplicadas. 10](#_Toc80637325)

[3.9 – Salvar o DataFrame em formato CSV. 10](#_Toc80637326)

[4 CONCLUSÃO 10](#_Toc80637327)

# INTRODUÇÃO

Porter (1998) define Inteligência Competitiva como coleta de informações sobre os movimentos de concorrentes e sobre seus padrões de ação estratégica e de crescimento. A Inteligência pode estar associada a um sistema que utiliza informações para avaliar os recursos organizacionais a fim de implementar estratégias para lidar com o seu complexo e, em constante mutação, com o seu ambiente cultural (Cruz, 2016; Muhammad et al., 2017).

Este projeto de conclusão da disciplina de Localização e Utilização da Informação tem por finalidade demonstrar as possibilidades de extração de informações diretamente de sites públicos utilizando dois dos métodos mais eficazes e conhecidos, API ou Scraping, técnicas que visam fazer requisições de dados diretamente das fontes, para posterior inferências no intuito de trazer algum tipo de vantagem competitiva para as empresas, seja melhoria de preços, retenção de clientes, marketing, dentre outros.

Inteligência Competitiva significa um processo sistemático com o propósito de coletar e analisar informações sobre os concorrentes e o ambiente econômico e sociopolítico das empresas. O seu maior objetivo é utilizar fontes distintas de recursos a fim de aumentar a competitividade da organização e fazer decrescer esse mesmo fator em rivais (Cruz, 2016).

## API

API ***(Application Programming Interface)*** ou em português, Interface de Programação de Aplicações, é um recurso intermediário entre softwares, trata-se de um conjunto de rotinas e padrões estabelecidos e documentados por uma aplicação que permite um software fazer referência a outro software distinto. Em resumo, um API entrega buscas solicitadas por um provedor e então responde essas informações aos solicitantes.

## WEB SCRAPPING

É um processo que faz analogia a uma “raspagem” de dados vindos de fontes públicas através de script ou programas, e então armazenar esses dados coletados em algum dispositivo.

Este processo funciona em duas partes, começando com o WEB CRAWLER, que é uma busca mais robusta e conhecido como rastreador, que faz um apanhado geral do tema principal a ser coletado, e então entra o WEB SCRAPING, uma busca mais detalhada e fina da informação desejada.

# OBJETIVO

Escolher um tema, localizar a fonte desejada, utilizar uma das técnicas de extração e aquisição que foi estudada na disciplina, armazenar os dados em um DataFrame, salvar em formato JSON ou CSV, e então realizar algumas análises exploratórias.

Foi definido que a Localização, Extração e Utilização dos dados será através do ***Spotify***, um dos maiores provedores de música por streaming conhecidos mundialmente.

A ideia para criação desse projeto, foi utilizar um material de referência, pesquisado no Kaggle, onde o script em Python, possibilita além da extração de dados dos artistas, definir as ***FEATURES (características) de cada faixa*** produzida por esses artistas, bem como a popularidade alcançada por cada música.

Sendo assim, essas informações criam a possibilidade de utilizar esses dados de diversas maneiras, seja criando Lista de Reprodução direcionadas, ou seja, visualizando qual das faixas de cada artista se encaixa em um melhor horário do dia, de forma a “orientar” o humor e bem-estar de seus usuários.

Também é possível desde que o usuário permita, acessar seus dados de utilização e criar Playlists direcionadas de acordo com os estilos e artistas mais escutados pelo usuário.

## - Extrair os dados (API ou Scraping).

Foi escolhida a técnica de **API**, já que a empresa disponibiliza o ambiente de desenvolvedor, onde é possível criar uma [WEB API](https://developer.spotify.com/documentation/web-api/) para fazer tanto Scraping como interconectar o **Spotify** a um outro aplicativo específico que utiliza de recursos musicais.

## 2.2 - Alocar essas informações em um DataFrame.

Após a **extração dos dados**, foi possível através de um **script em Python**, alocar todos os dados escolhidos em um DataFrame para posterior inferências .

## - Executar a operação de ordenação no DataFrame em um atributo à sua escolha.

Os **DataFrame** criados foram **salvos em formato CSV** onde cada um desses DataFrame levou o nome do artista selecionados a futuras inferências.

## - Executar a operação de filtragem no DataFrame à sua escolha.

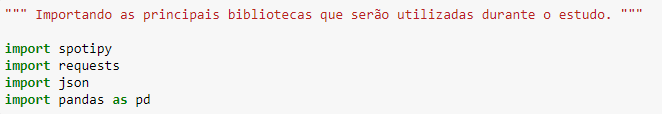
As filtragens feitas foi para definir quais músicas são as mais escutadas por característica e utilizar dessa informação para possibilidade da criação de Lista de Reprodução direcionada.

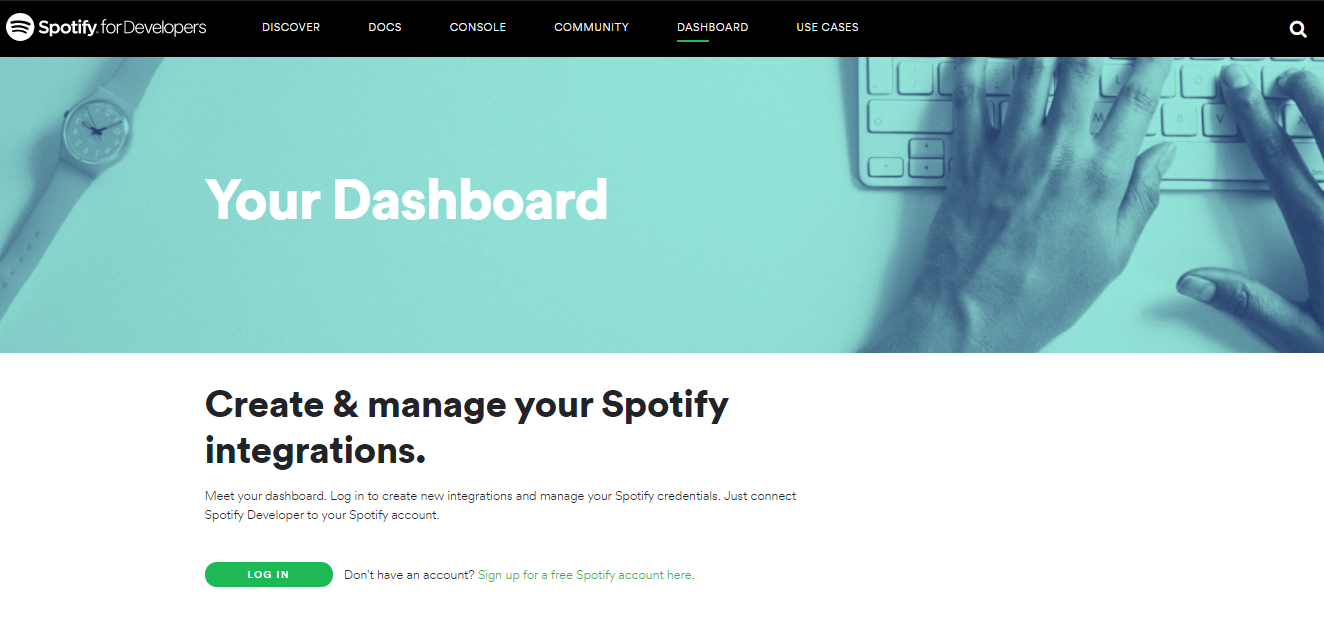
# DESENVOLVIMENTO

A seguir serão desmembradas as etapas para o desenvolvimento de Localização e Armazenamento de dados através da tecnologia de API, que foi a escolhida para o desenvolvimento desse projeto.

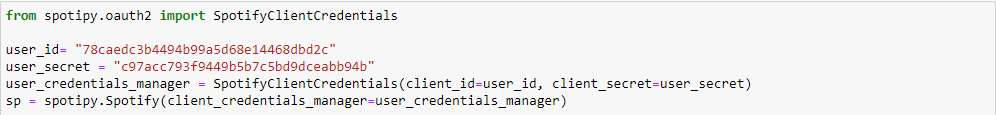
## – Carregar Bibliotecas e Cadastramento da WEB API e obtenção do ID e KEY.

A primeira etapa é carregar as bibliotecas e o cadastramento do usuário como desenvolvedor dentro da plataforma do próprio Spotify, essa etapa pode ser feita através deste [link](https://developer.spotify.com/dashboard/login).

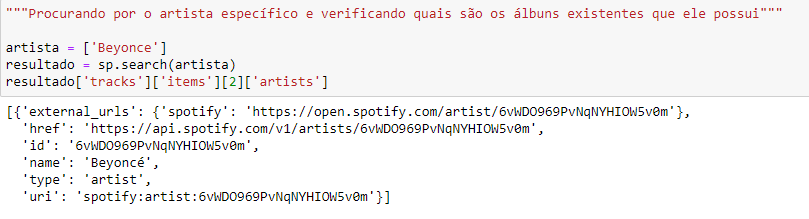




Após o cadastramento e aquisição do ID e Key de usuário o código abaixo pode ser utilizado para se comunicar com o Spotify para coleta de dados diretamente da plataforma.

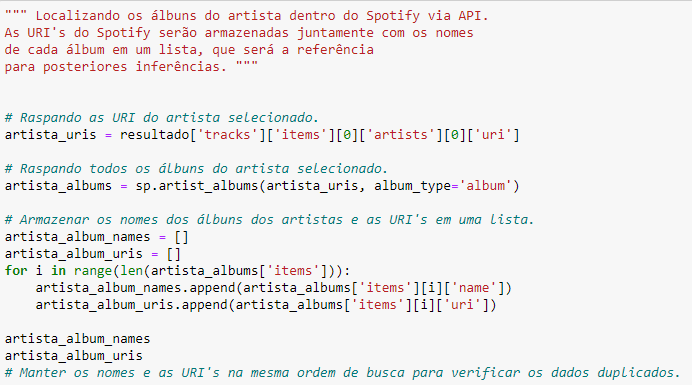


## – Definir artista e realizar a raspagem com a tecnologia API.

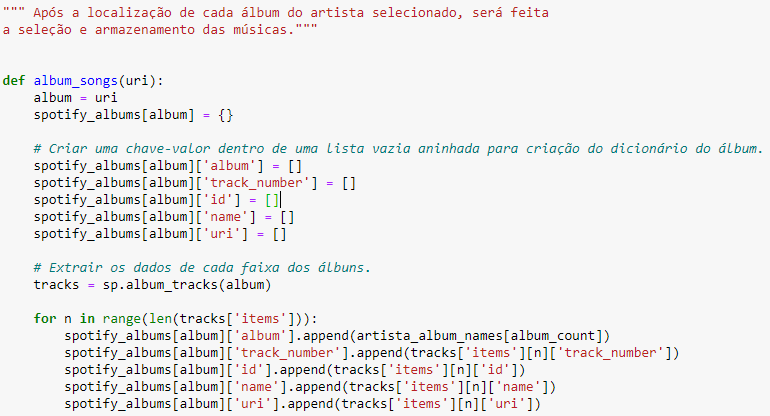


Ao realizar a chamada dos dados através da API do Spotify, essas são as respostas vindas da plataforma.

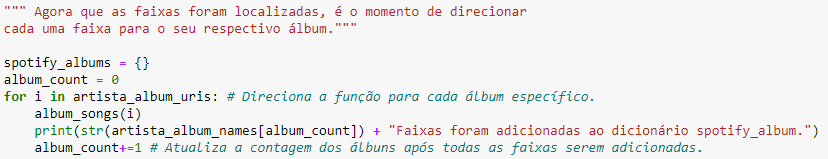
## – Raspar as URI’s de todos os álbuns dentro do Spotify.



## Selecionar todas as músicas de cada álbum.



## – Adicionar cada música ao dicionário do álbum correspondente.



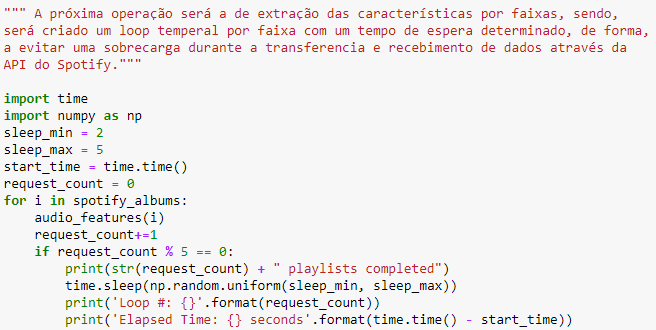
## 3.6 – Extrair e exibir as características de cada faixa.

Cada faixa possui 10 características próprias que já estão inseridas na própria API do Spotify, o que permite que os desenvolvedores trabalhem ideias a partir dessas Features próprias de cada faixa musical do artista, uma das formas mais comuns de iteração com outras aplicações, é usar essas features como base para sugestão de acordo com a personalidade musical de cada usuário.



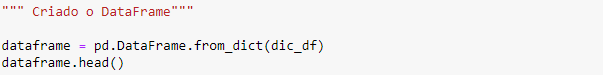
* **ACOUSTICNESS**: Este valor descreve a quão acústica é uma música. Uma pontuação de 1,0 significa que é mais provável que a música seja acústica.
* **DANCEABILITY**: “A capacidade de dança descreve o quão adequada uma faixa é para dançar com base em uma combinação de elementos musicais, incluindo tempo, estabilidade do ritmo, força da batida e regularidade geral. Um valor de 0,0 é menos dançável e 1,0 é mais dançante”
* **ENERGY**: “Representa uma medida perceptual de intensidade e atividade. Normalmente, faixas energéticas parecem rápidas, altas e barulhentas”.
* **INSTRUMENTALNESS**: Este valor representa o número de vocais da música. Quanto mais próximo estiver de 1.0, mais instrumental será a música.
* **LIVENESS**: Este valor descreve a probabilidade de a música ter sido gravada com um público ao vivo. De acordo com a documentação oficial “um valor acima de 0,8 fornece grande probabilidade de a faixa estar ao vivo”.
* **LOUDNESS**: “É a percepção subjetiva da pressão sonora. Mais formalmente, é definido como, "Aquele atributo da sensação auditiva em termos dos quais os sons podem ser ordenados em uma escala que vai do baixo ao alto".
* **SPEECHINESS**: “A fala detecta a presença de palavras faladas em uma faixa”. Se a fala de uma música estiver acima de 0,66, provavelmente é composta de palavras faladas, uma pontuação entre 0,33 e 0,66 é uma música que pode conter música e palavras e uma pontuação abaixo de 0,33 significa que a música não tem fala.
* **VALENCE**: “Uma medida de 0,0 a 1,0 que descreve a positividade musical transmitida por uma faixa. Faixas com alta valência soam mais positivas (por exemplo, feliz, alegre, eufórico), enquanto faixas com baixa valência soam mais negativas (por exemplo, triste, deprimido, com raiva)”.
* **POPULARITY**: Popularidade e representa a quantidade de vezes que essa música é escutada pelo usuário.

## – Recursos temporal organizador para tornar a operação de extração das características de cada faixa mais dinâmica e rápida.

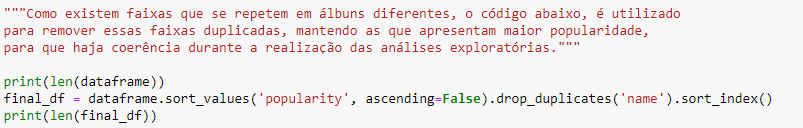


## – Criando o DataFrame para o artista, álbum, faixa, features e popularidade.

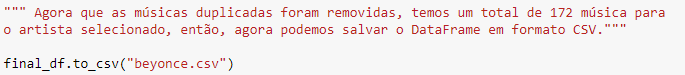




## – Eliminar as músicas duplicadas.



## 3.9 – Salvar o DataFrame em formato CSV.



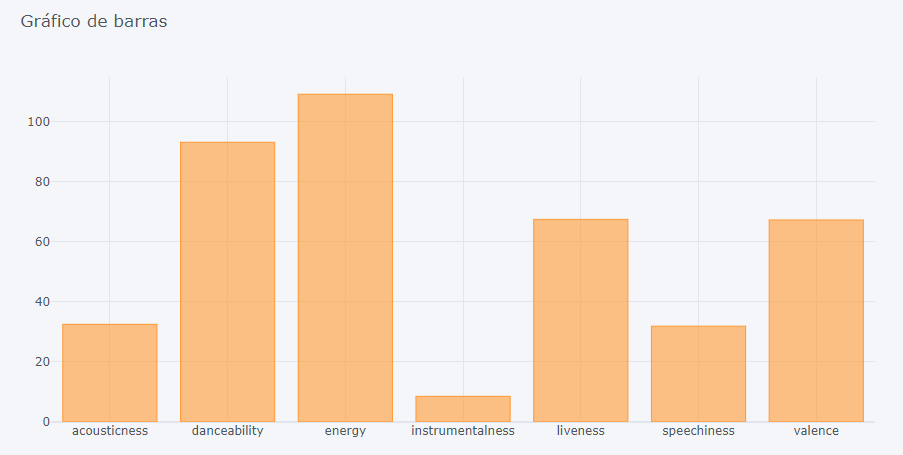
# CONCLUSÃO

A partir do entendimento de como funciona a tecnologia de API e de WEB Scraping, é possível entender que essas técnicas vão além de uma simples tarefa mecânica de coleta de organização de dados.

A Inteligência Competitiva, segundo Tarapanoff (2001), é uma nova síntese teórica no tratamento da informação para a tomada de decisão. Compreende uma metodologia que permite o monitoramento informacional da ambiência e, quando sistematizada e analisada, possibilita a tomada de decisão.

Essas técnicas são peças fundamentais na construção do que hoje chamamos de Inteligência Competitiva, um ramo em que todas as empresas estão se especializando tanto para conhecer mais a fundo os seus concorrentes, como também para entender melhor como é o comportamento de seus clientes e até mesmo parceiros de negócios.

De acordo com Choo (2002), os tomadores de decisão não são apresentados automaticamente aos problemas que eles precisam resolver e, tampouco, às soluções alternativas que podem escolher. Eles devem identificar os problemas, buscar soluções e desenvolver métodos para gerar e avaliar as alternativas.



O gráfico acima foi construído a partir da aquisição de dados de um artista específico pela API do Spotify, esse gráfico representa as características e o somatório dessas com relação ao artista estudado, dessa forma, é possível direcionar esse artista para público que possuam em seu histórico artistas com característica musical semelhante, bem como sugerir artistas para momentos específicos do dia e da semana.

Hoje em dia contamos com a opção de escutar álbuns independentes criados pelo próprio Spotify para uma Sexta-Feira a noite, uma manhã de Domingo, um álbum para estudar, ou até mesmo álbuns separados por categoria de estilo.