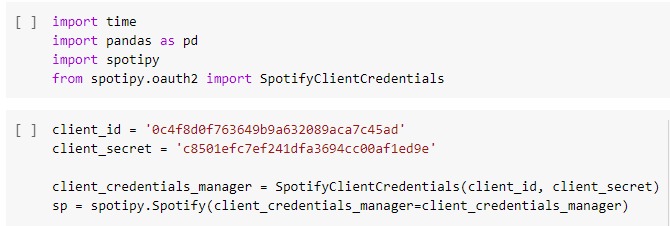
**Análise em python: Top hits do Spotify dos anos 2000 a 2021. O que existe em comum entre as músicas mais ouvidas de cada geração?**

****

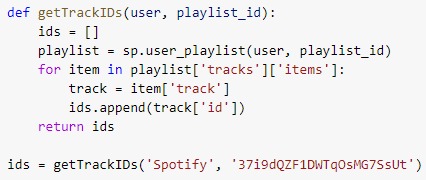
Você ouve música todo dia? Consome muito o Spotify? Nós, do grupo 2 de trainees 2022 também!

Nesse artigo, abordaremos a análise das músicas mais tocadas no spotify mundial desde o ano 2000 até o ano de 2021,(dado que a lista das músicas mais tocadas de 2022 ainda não tenha sido finalizada, não à usaremos) mostrando o que há em comum, ou não, entre os top hits de cada ano, o que mudou em quesitos de preferências musicais, de uma década para a outra, será que outras redes sociais podem impactar nas antelações musicais das pessoas? Realizaremos tudo isso usando o nosso querido python, trabalhando com API’s doo spotify, webscraping e utilizando pandas para as nossas análises. Prepare os fones de ouvido e o nosso aplicativo tão querido e venha acompanhar esse artigo com a gente!

Para começar nosso projeto, primeiro importamos algumas bibliotecas cruciais para a realização da análise, dentre elas a “spotipy”, para podermos entrar na API do spotify. Porém, para conseguirmos ter liberação total da API do spotify tivemos que pegar credenciais específicas do spotify, representadas pelas duas variáveis abaixo, fato que só conseguimos realizar após criarmos uma conta no spotify como **desenvolvedor**, somente assim.



Logo após, criamos uma função básica com o propósito de capturar o código específico de cada música, (todas músicas em qualquer plataforma de transmissão tem seu código específico, a análise pelo software python tem que ser feita com ele) tendo duas variáveis em seu interior, sendo o “ user” a nossa conta no spotify de onde retiramos nossas playlists, e a outra variável representando os códigos de cada playlist. Passando por todas as músicas de cada playlist e armazenando seu código.



Poderíamos ter pego todas as 22 playlists a mão? Sim, por ser uma quantidade baixa se torna viável, mas não seríamos nós mesmo se deixássemos esse processo passar sem automatizá-lo, a graça está em mostrar que com 22 playlists ou com 1000 a usabilidade e eficiência se torna a mesma. Para tal automatização utilizamos o selenium e fizemos o webscraping no spotify, passando por todas as playlists definidas dentro do “range”, pegando seus “id’s” ou códigos específicos e armazenando todos numa lista.



Por conseguinte, criamos uma função da qual permite relacionar os id’s coletados com os nomes já fixos do spotify, que transformaremos em colunas no data frame. Como fizemos: criamos a variável “meta”, contendo todos os id’s de todas as músicas, logo em seguida criamos a “features”, que traz diversas características das músicas do spotify, como artista, tempo, data de lançamento, etc. Dessa forma, construímos várias variáveis das quais pertenciam as características de features e, relacionamos cada uma com a variável meta, para construirmos as colunas devidamente indexadas com seus respectivos id’s.

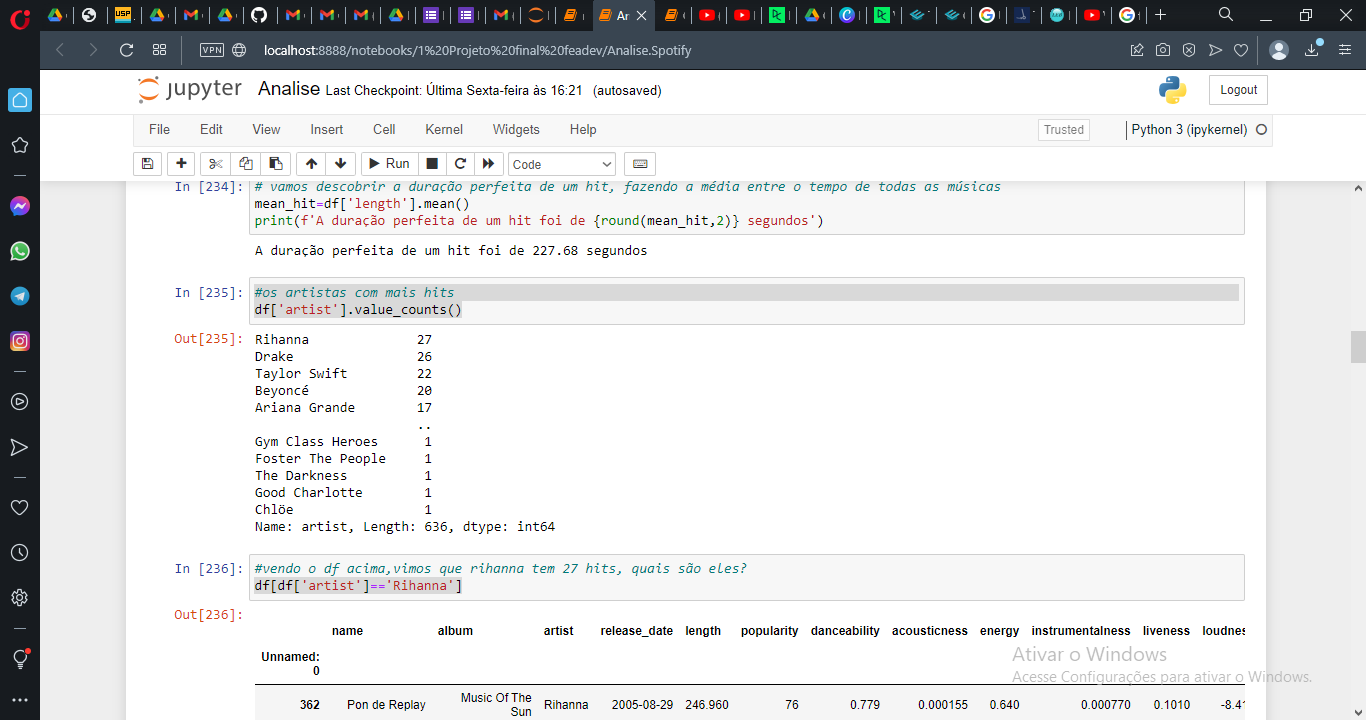
Cria-se uma track, com o nome de todas essas variáveis para finalmente montarmos o nosso data frame.



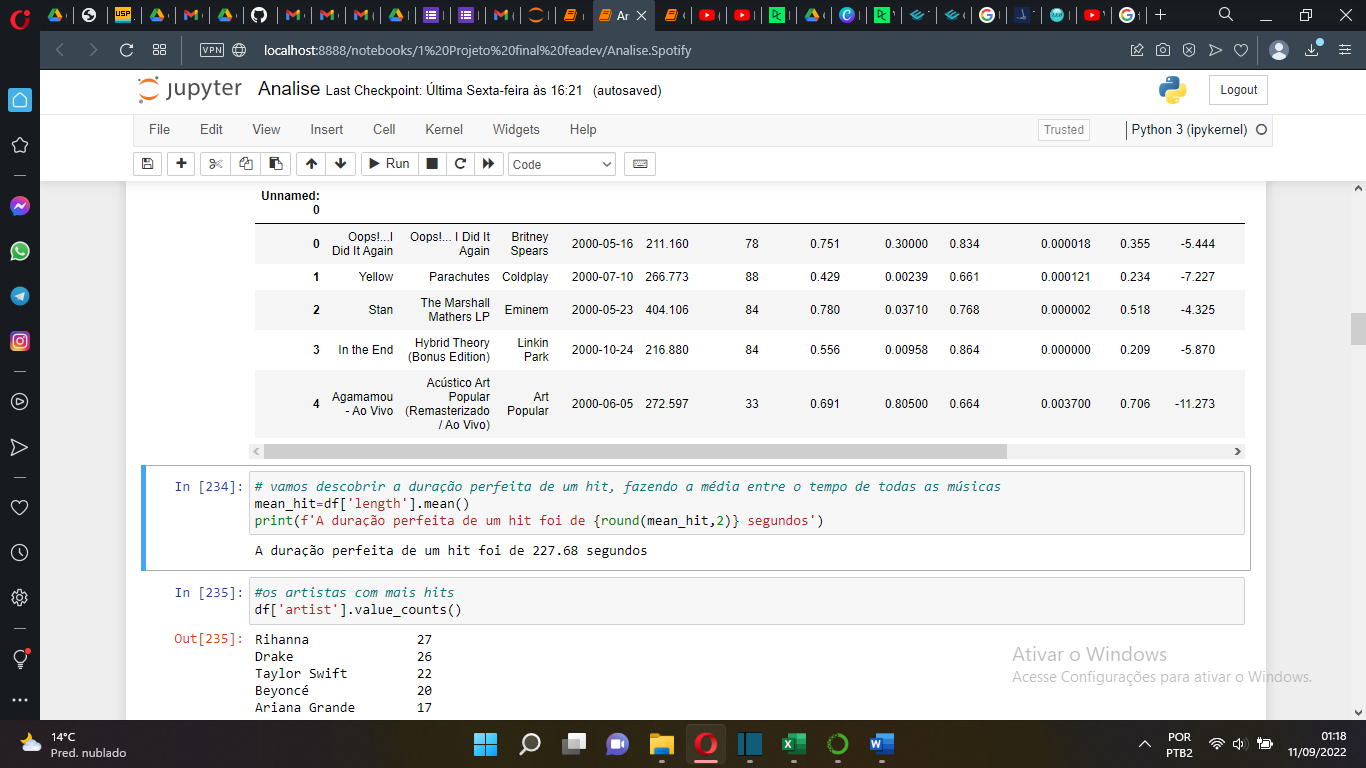
Por fim, após colocarmos em ordem o restante dos dados, chamamos a biblioteca pandas com o “pd” e transformamos nossa variável track em um data frame, passando seus valores internos como colunas.



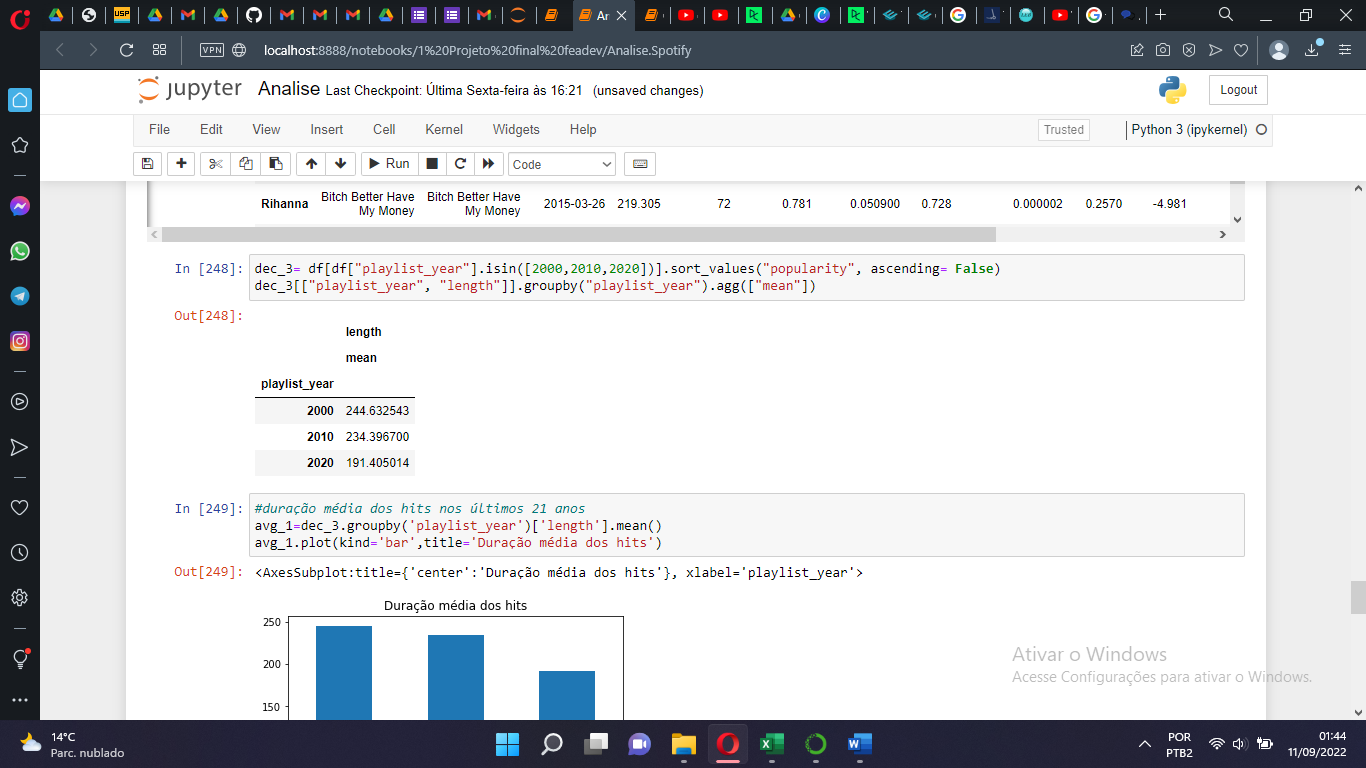
Quanto a análise de dados em si, conseguimos algumas informações curiosas, como saber que a cantora Rihanna foi a artista que mais teve hits nos últimos 21 anos, totalizando 27 músicas de sucesso. Para isso, usamos a função “value\_counts”, a qual mostra quantas vezes os elementos da coluna “artist” apareceram no data frame.



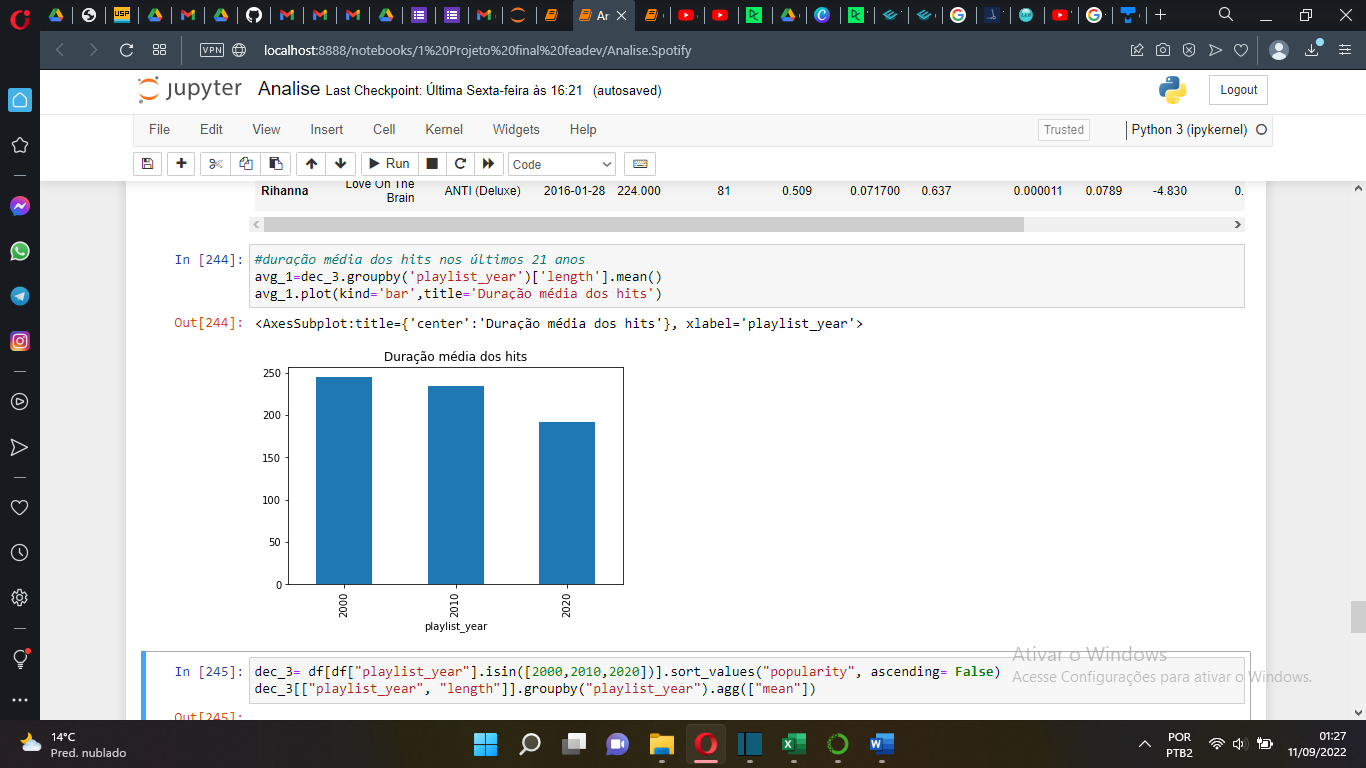
Além disso, vimos que em média, um hit possui 3 minutos e 47 segundos, fazendo a média entre todas as músicas tiradas das playlists desde os anos 2000, utilizando a função “mean()”.



Outro fato importante, foi a constatação de que as músicas do ano de 2020 ficaram mais curtas em comparação com as de duas décadas anteriores. Para chegar nesse resultado, primeiramente, com a ajuda da função “isin”, nós fizemos um novo subconjunto do data frame, chamado “dec\_3”, o qual só possui os hits dos anos 2000,2010 e 2020. Assim, usando o “groupby()” e o “agg()”, filtramos nesse novo subconjunto, levando em consideração a duração média das músicas, que estão na coluna “length”.

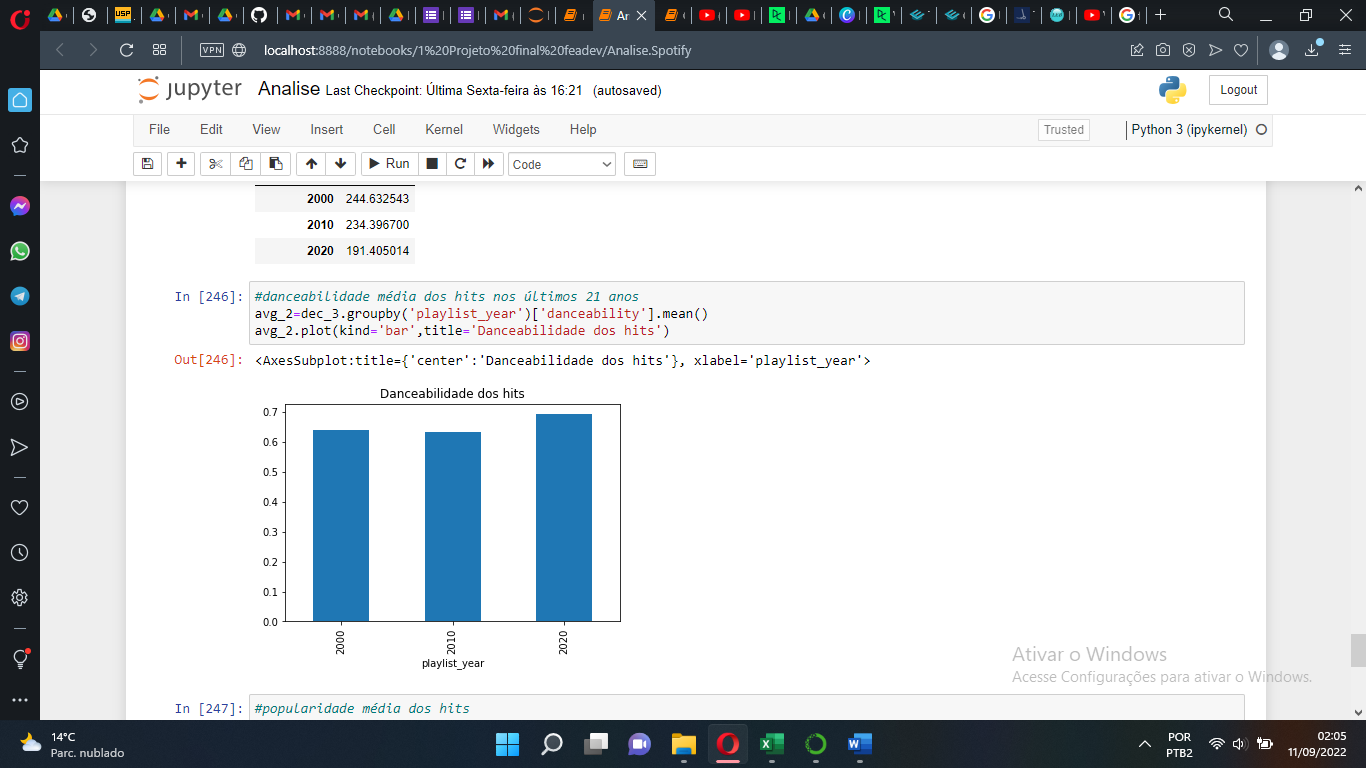


Depois, para melhorar na visualização dos dados, usamos a função “plot()” que plotou um gráfico de barras da diferença do tempo das músicas nas playlists de 2000,2010 e 2020.



Essa diminuição da duração dos hits, se deve ao fenômeno da “audição ansiosa”, ou seja, as pessoas querem músicas mais rápidas e mais objetivas, já que se elas não gostam da faixa, em segundos elas vão trocar e ir para a próxima. Assim como acontece nos stories, que possuem 15 segundos de duração e nos tiktoks, que têm 30 segundos. Desse modo, os cantores estão investindo em músicas com menor tempo para aumentar as chances de emplacar um hit.

Ademais, percebemos também que os hits em 2020 estão mais dançáveis que as últimas décadas. Para isso, usamos o recorte do data frame que já havíamos usado anteriormente dos anos de 2000,2010 e 2020, com a única diferença de que na hora de plotar o gráfico, usamos a coluna “danceability”.



Com isso, esse resultado também pode ser associado a necessidade dos artistas lançarem músicas dançáveis que façam sucesso no tiktok, já que essa rede social é um ótimo jeito para divulgar o trabalho deles.