

## 5.6 - Conectando SQL com o pandas (parte 1)

00:00:00:00 - 00:00:01:17

Olá, pessoal. Bem-vindes a mais uma aula. Bom, gente, nós já aprendemos como filtrar os dados que precisamos do banco de dados, mas nós precisamos tirar os dados de lá para juntar com outras análises. Nós poderíamos simplesmente salvar o resultado da query em um arquivo lá no DBeaver, mas vamos aprender uma maneira mais prática. Vamos aprender a fazer nossa query direto no Python e a conectar direto no Pandas para, assim, juntar com os outros dados que temos. Bora lá. Bom pessoal, normalmente, para fazer a conexão com um banco de dados através do Python, nós utilizamos três informações básicas. Vou para cá. Aqui. O host aqui, que é o endereço na internet onde está o banco, normalmente um IP, ou se ele estiver no nosso computador local, usamos endereço como localhost, como está aqui nesse exemplo. Temos que passar o user, que é o nome do usuário, e o password, que é a senha. No exemplo, temos um trecho de código de como faríamos se estivéssemos usando o banco MySQL aqui. O módulo mysql ponto connector, esse módulozinho aqui que está fazendo o import dele, é uma biblioteca em Python usada para conectar com um banco de dados MySQL a partir do Python.

00:00:01:17 - 00:00:02:18

Na linha abaixo, é criada uma conexão com o banco de dados usando ele, usando esse método connect e passando os três parâmetros que falamos. Mas o nosso banco de dados foi criado com o SQLite, que é um banco de dados muito mais leve, normalmente usado nos aplicativos dos nossos celulares. O SQLite é um banco de dados embutido, o que significa que o arquivo do banco de dados é armazenado localmente no sistema de arquivos. Isso torna altamente portátil, permitindo que a gente mova facilmente o banco de dados entre diferentes plataformas. Então, nós podemos pegar o arquivo do nosso banco de dados e subir lá para a nossa pasta no Drive e fazer as nossas consultas diretamente no notebook do Google Colab. Vamos começar encontrando aquele arquivo que criamos lá na aula quando a gente criou o banco de dados. Vocês lembram dele? A gente salvou o nome do arquivo como Status Brasil. No meu caso, eu lembro que eu salvei lá na pasta de documentos porque achei que era melhor para eu poder encontrar ele.

00:00:02:18 - 00:00:03:34

Primeiro eu vou aqui, só salvar tudo aqui, só para a gente ter certeza que está tudo certinho aqui no DBeaver. Salvei tudo, aí eu vou abrir aqui a minha pasta de documentos. E aqui está: Status Brasil. E eu vou arrastar esse Status Brasil lá para a minha pasta no Drive, onde eu estou com todas as minhas coisinhas do curso. Estou com o meu notebook aqui "análise de dados" e tal, então eu vou arrastar daqui para lá. Vou diminuir essa telinha para ficar mais fácil, aí eu pego daqui e arrasto aqui para baixo. Bom, assim ele já vai subir aqui para o Drive e já vai aparecer aqui para mim disponível. E aí eu vou lá para o meu notebook de análise de dados. Eu vou criar um índice aqui chamado conectando SQL com pandas. Adicionar uma nova célula de texto: conectando SQL com pandas. Bom que fica tudo certinho e a partir daqui é tudo relacionado ao SQL, ok? e primeiro nós vamos importar a biblioteca sqlite3, que assim como a biblioteca MySQL, serve para conectar, interagir e manipular os bancos de dados, porém, apenas com bancos do tipo SQLite, ok? Então: import (tudo minúsculo) sqlite3, beleza?

00:00:03:34 - 00:00:04:54

Vou executar. Agora, vamos criar uma conexão com o banco de dados através da função connect. Dentro da função, nós vamos passar o caminho do arquivo que a gente acabou de adicionar no drive. Então, a gente vai colocar sqlite3 ponto connect. E aí, dentro dos parênteses, eu vou colocar o arquivo. Não sei se vocês lembram, mas a gente vai aqui na aba esquerda, clica aqui nessa pastinha, já vai estar conectado com o drive porque eu executei tudo desde o início, todos os códigos que a gente fez até agora eu já executei, e lá no início tem aquela conexão que a gente cria com o drive. Então, no meu caso, já está conectado. Vou nessa parte de My Drive, vou lá na pasta PrograMaria, que é onde estão as coisinhas todas, e aqui eu procuro o meu arquivo Status Brasil, que eu acabei de subir. Está aqui. Clico nesses três pontinhos e vou em copiar caminho. Beleza. E aí eu colo aqui dentro dos parênteses, coloco a aspas e colo aqui o caminho. Vou fechar aqui já para ficar mais fácil. Beleza, e essa conexão aqui eu vou atribuir a uma variável chamada conexão.

00:00:04:54 - 00:00:06:04

Muito criativa. Então aqui, no início: conexão. Beleza, essa variável será usada para referenciar a conexão com o banco. Basicamente temos tudo o que a gente precisa. Agora vamos fazer uma consulta simples para saber se está tudo certinho. Primeiro vou executar para colocar essa conexão aí de pé. Beleza. Lembra daquela consulta que a gente fez sobre o Itaqua? Vamos fazer ela aqui de novo. Vamos criar uma variável chamada query que vai receber a nossa consulta. E como já vimos essa consulta aqui no banco de dados, podemos só ir lá no DBeaver, copiar e colar aqui. Então, eu vou chamar aqui, criar essa variável query, vou lá no DBeaver, onde tem as minhas queries aqui que eu acabei de fazer, da aula passada, e vou pegar essa daqui de Itaqua. Vou voltar lá para a query, abrir as aspas e colocar aqui dentro a minha consulta. Ao invés de consultar o like, eu vou colocar o igual para a gente poder pegar só o resultado de Itaqua. Itaquaquetuba. Beleza.

00:00:06:04 - 00:00:07:18

Olha, se vocês repararem, ficou errado aqui, ficou estranho. Isso porque eu coloquei Itaquaquetuba entre aspas simples e coloquei a frase toda entre aspas simples também. Então tá meio que dando uma convergência ali de aspas. Pra não dar isso, eu vou colocar aspas duplas aqui do lado de fora. Aspas duplas e pronto, agora tá certinho. Eu posso até executar essa query aqui em uma linha de código abaixo, chamar a query de novo pra gente poder visualizar aí a nossa frase. Beleza. Selecionar tudo da tabela municípios brasileiros, onde cidade é igual a Itaqua, ok? Então, para poder fazer a conexão com o SQL, conseguir fazer essa consulta, a gente vai usar um método do Pandas que é o read SQL, que é pd ponto read sql, e os parênteses aqui. Dentro dos parênteses, vamos colocar dois parâmetros. O primeiro é a nossa query e o segundo é um parâmetro chamado "com", que vai ser igual à nossa variável de conexão. Então, aqui a gente manda primeiro a query e depois o parâmetro "com" que vai ser igual à conexão. Beleza?

00:00:07:18 - 00:00:08:14

E aí a gente pode executar e observar o resultado. E olha só, é o mesmo resultado que a gente recebeu quando fez essa consulta lá no DBeaver. Agora você pode estar aí se perguntando o que a gente vai fazer com esse banco de dados, qual que vai ser a análise e tal. Mas pense comigo, nós temos aqui nesse banco uma entidade, uma tabela, com colunas sobre índice de educação, renda, etc, de cada município do Brasil, correto? Esses municípios pertencem a estados, né? E nos dados que analisamos no último módulo, com a pesquisa lá das pessoas que trabalham na área de TI, nós temos uma coluna de estados. Vejam só. E se a gente juntasse essas duas coisas e conseguisse responder à seguinte pergunta: será que existe correlação entre índice de renda e o salário? Para sabermos se os estados com renda maior oferecem salários maiores para os profissionais de TI, por exemplo.