

## 5.6 - Conectando SQL com o pandas (parte 2)

00:00:00:00 - 00:00:01:02

Vamos por partes. Primeiro, vamos relembrar como estão as nossas tabelas no banco de dados. Lá no DBeaver, se a gente for aqui em Status Brasil, clicar duas vezes aqui na opção de tabela, a gente clica duas vezes em tabela, e aqui no ER Diagrama a gente consegue ver os diagramas, né? Nós queremos o índice de renda que está na entidade município status. Então, aqui, a gente quer esse renda aqui, que é o índice de renda. Mas precisamos da coluna de estado que está na entidade de municípios brasileiros. Vocês lembram que nessa coluna de município status, o município ID são só números, e aí a gente precisava juntar com a tabela de municípios brasileiros para ter o nome da cidade, o nome do estado, região e tudo mais? Então, a gente pode fazer um join, que a gente já aprendeu na última aula para juntar essas duas tabelas. Inclusive, já fizemos esse join aí, a gente pode só copiar para poder poupar em um pouco de tempo. Aliás, para ficar mais fácil ainda, a gente pode colocar essa consulta lá na nossa variável chamada query, que aí fica mais fácil também de executar tudo.

00:00:01:02 - 00:00:02:31

Então, vamos pegar aqui... Olha... Onde a gente fez esse join? Aqui, a gente fez esse join nessa parte. Vamos copiar. Voltar lá para o nosso notebook. Criar aqui mais uma linha de código e criar aí de novo a query e colocar aqui dentro toda a nossa query que a gente copiou do DBeaver. A primeira coisa foi: eu copieei lá o DBeaver, tinha esses dois tracinhos aqui de comentário. Já vou tirar ele aqui porque a gente não precisa mais, vai ser um código mesmo, né? Segunda coisa, eu vou dar vários tabs aqui para poder ficar na mesma linha aqui, só para ficar melhor de visualizar, ok? Terceira coisa é que isso aqui tudo precisa estar entre aspas, né? E aí se eu colocar aspas aqui e aspas no final aqui embaixo, não vai dar certo, tá vendo que não pegou tudo? Quando a gente quer que seja comentário, ou seja texto, na verdade, nessas duas linhas, a gente usa três aspas seguidas. Assim. Tá vendo que eu coloquei as três aspas aqui embaixo já ficou vermelhinho também? Aí eu tô querendo dizer que tudo que tá depois das aspas, dessas três aspas, é texto. Tanto essa linha, quanto a de baixo. Inclusive, se eu criar uma linha abaixo dessa daqui, tipo, se eu levar isso aqui lá pra baixo, tá vendo?

00:00:02:31 - 00:00:03:44

Continua vermelho, olha. Porque tudo é como se fosse um texto também, tudo junto. E para fechar esse texto, eu vou lá no final e coloco três aspas. Opa, coloquei quatro. Três aspas. Pronto, fechei. Se eu colocar algum texto aqui embaixo, já não vai ficar vermelhinho, já não é texto mais, porque eu fechei lá com as três aspas. Então, ok. Temos aí a nossa query que faz o join. Bom, eu peguei aqui a query para poder facilitar, mas nessa query a gente estava pegando cidade e pegando a população residente dos municípios brasileiros. Agora eu não quero cidade, né? Eu quero estado. Então, eu vou colocar aqui: estado. E eu não quero população residente, eu quero renda, né? Então, eu vou colocar aqui: renda. Além disso, eu vou pegar a média da renda por estado, né? Eu vou agrupar tudo aqui pela média da renda. Então, eu vou usar um comando chamado AVG que calcula a média, como se fosse o min que a gente já viu anteriormente do NumPy e do Pandas, beleza? Então, beleza, aqui tem front, tem o join que a gente precisava, ótimo.

00:00:03:44 - 00:00:05:00

Mas a gente precisa fazer só uma coisinha que é o group by por estado no final. Então aqui no finalzinho, eu já fiz o join, eu vou dar um enter, tá dentro das três aspas, então tá tudo certo, e vou colocar um group by. E aí eu posso copiar isso aqui, que vai ser exatamente municípios brasileiros e estados. E aí fica assim a nossa query. Voltando tudo aqui para a gente poder entender o que a gente está fazendo. A gente está selecionando o estado, a média da renda por estado da tabela de municípios brasileiros. E aí a gente tá fazendo um join, porque o estado está numa tabela e a renda está em outra. Então a gente está fazendo o join de duas tabelas e no final a gente tá agrupando por estado. Por que a gente tá agrupando por estado? Pra não ficar aquele tanto de linha com o mesmo estado, né? Minas, Minas, Minas, Minas... Então a gente tá agrupando tudo por estado e aí pega a média do índice de renda. Então, ok. Vamos executar isso aqui. E aí aqui embaixo a gente pode copiar isso aqui do jeito que está: read sql e colar aqui embaixo porque vai ser exatamente isso. A gente precisa executar esse negócio de query para atualizar porque agora a nossa query é essa. E aqui embaixo a gente executa de novo e pronto, olha o nosso resultado.

00:00:05:00 - 00:00:06:16

A gente tem aqui na primeira coluna o estado e na segunda coluna a média da renda por cada estado. Mas aqui nós temos todos os estados do Brasil, porém lá nos nossos dados da pesquisa, do Data Hackers, não temos todos os estados. Para que nossa consulta no banco contenha apenas os estados que tem lá na pesquisa do Data Hackers, nós podemos listar os estados da pesquisa e fazer a consulta por cima dessa lista. Vamos por etapas. Primeiro vamos ler o nosso arquivo e listar os estados. Lembra que salvamos o nosso arquivo com formato CSV? A gente pode fazer a leitura desse arquivo de novo usando o Pandas. Então, vamos lá. Vamos chamar de novo de dados, né? pd ponto read csv porque o nosso arquivo é em formato CSV. E aí a gente pode vir aqui no cantinho e procurar lá o nosso arquivo que a gente salvou: análise dados ponto csv, dentro da pasta que eu salvei, da PrograMaria e tudo mais. Vou aqui nos três pontinhos, copiar caminho, fechar aqui já, aspas, e dentro das aspas eu coloco o caminho do arquivo. Ok. E vou fazer a leitura dele. Agora vamos listar os estados únicos que temos nessa coluna.

00:00:06:16 - 00:00:07:34

Na coluna de estado, né? Eu não lembro qual que era o nome da coluna de estados, então eu vou colocar aqui: dados ponto columns, que vai listar aí todas as colunas para mim. E aí eu tenho a coluna de UF onde mora, que é a coluna de estados. Tanto que se a gente for aqui e colocar: dados, colchetes e UF onde mora, vai ter aqui todas as entradas de estado. Bom, agora vamos listar os estados únicos. Aqui está os estados, mas está repetido, né? Eu preciso da lista dos estados únicos dessa coluna aqui de UF onde mora. Para poder pegar os estados únicos, eu posso colocar um "ponto unique", que é uma função que pega tudo que é único em uma única coluna. Então, aqui eu tenho todos os estados. Todos não, né? A gente pode perceber, por exemplo, que não tem um Acre, enfim, então por isso que a gente está fazendo isso, para não precisar ir lá no banco de dados e filtrar dados que a gente não precisa. Aí a gente pega só desses estados que estão aqui na nossa base de dados. Só que aqui não está em formato de lista, está em formato de array. Para ficar em formato de lista mesmo, a gente usa a função de list. O list, tudo que estiver entre parênteses aqui, dentro dos parênteses do list, vai ser transformado em uma lista.

00:00:07:34 - 00:00:09:08

Se eu executar aqui, uma listinha de valores únicos. Vamos criar aí uma variável que vai receber esse resultado: lista estados. Pronto. Voltando lá para a nossa query, vamos adicionar uma linha depois do join. Vamos voltar lá na nossa query. Aqui a gente tem o join. A gente pode adicionar depois do join uma condicional. A condicional é que só quero os resultados dos estados que estão na lista estados, que estão dentro daquela lista que a gente criou. Então, a gente pode usar o where, que a gente está falando de condicional. Então: where estado esteja in, ou seja, onde tem os estados na lista X. Então: join tal aqui (join municípios, município ID, tal) where aqui, e aí where estado, ou seja, município brasileiro estados in lista estados. Seria lindo demais se a gente pudesse só vir aqui e colocar: lista estados. Mas se a gente executar vai dar erro, pois a lista estado é uma lista. Precisamos juntar essa lista de string, de lista com textos, com a string completa da nossa query. Então, vamos fazer uma pausa aqui antes para entender alguns comportamentos de string.