

5.6 - Conectando SQL com o pandas (parte 4)

00:00:00:00 - 00:00:00:58

E agora a gente pode voltar lá para a nossa query de SQL. Então no nosso SQL, quando a gente quiser preencher parâmetros, a gente não usa chaves igual a gente usa na string. A gente completa com um ponto de interrogação também. Então, o que a gente vai fazer? A gente vai primeiro passar a nossa lista de municípios, preencher com um ponto de interrogação. Então, a gente vai passar todo o tamanho da lista, mas todo preenchido com um ponto de interrogação. Então, a gente vai precisar da quantidade de pontos de interrogação que a gente tem de municípios para serem preenchidos. Para isso, o que a gente vai fazer? A gente vai usar a função join. E dentro dela, a gente vai criar uma lista de pontos de interrogações do tamanho da quantidade de municípios que a gente tem. Assim, ó... Voltar lá para a nossa query, ao invés aqui da lista de estados, a gente vai usar um ponto format, depois eu vou informar o que vem antes do ponto, tá? Mas vamos lá: ponto format... E aí dentro do parênteses a gente vai colocar que é separado por vírgula, ponto join, parênteses...

00:00:00:58 - 00:00:02:18

O que a gente vai preencher aqui dentro desse join, dentro do parênteses do join? A gente vai colocar que é uma lista. Uma lista de quê? De pontos de interrogação. Que vai ser o quê? Vai ser do tamanho da nossa lista de estados. Então, estou colocando que vai ser preenchido por interrogações durante o tamanho da lista de estados. Desse jeito, ok? Bom, eu coloquei aqui o format só pra poder mostrar pra vocês o que ele vai estar preenchendo, né? Que é esse espaço ali depois do in. Mas, como a gente já viu, o format ele vem depois da string. Então, eu vou tirar ele daqui, colocar depois a string o ponto format. Já tinha ponto, vou tirar. É um ponto só: ponto format. E lá depois do IN, eu vou colocar os parênteses e as chaves, ou seja, para indicar que esses valores do format aqui vão ser preenchido aqui dentro dessas chaves, ok? Então, beleza. Eu acho que eu não escrevi lista estados certo, deixa eu verificar se está o mesmo nome certinho. Acho que está, né? Vou voltar lá em cima. Lista estados, ok. Então, a gente pode executar essa célula e rodar o id SQL.

00:00:02:18 - 00:00:03:37

Bom, executando a gente pode vir aqui embaixo e ver que a query ficou um tanto de ponto de interrogação aqui, porque ficou estranho, não era assim que a gente estava querendo. Isso porque aqui faltou mais um parâmetro para ser enviado, que é chamado de parâmetro mesmo: params. E aqui a gente envia o lista estados. Ok. E aí a gente executa de novo. E agora sim, estão os estados aqui e a média da renda. Vocês podem perceber que, por exemplo, o Acre já não está aqui. Realmente a gente pegou só o que tava lá na lista de estados. Gente, eu fui criando aqui várias células de código e acabou que ficou uma ordem um pouco estranha, né? A gente está trabalhando com lista de estados aqui em cima, mas é lá embaixo que eu criei ela. Enfim, está fora da ordem. Então, eu vou levar esses dois itens lá pra baixo. Deixa eu ver se eu consigo selecionar. Isso, selecionei os dois usando o shift mesmo e vou colocar lá pra baixo. Pronto. Agora sim. Que aí a gente cria a lista de estados aqui, aqui está a query e aqui a gente faz a pesquisa. Bom, aqui está mais organizado, mas vamos dar uma voltada aqui nessa query.

00:00:03:37 - 00:00:04:57

Vamos printar ela. Print query, só para poder printar mais bonitinho. Então, vamos lá. A gente usou select, porque a gente quer os estados e a média da renda, ok? Estado está em uma tabela, renda está em outra. A gente quer isso da tabela de municípios brasileiros. Mas, como está uma em cada tabela, a gente faz o inner join. O inner join que vai pegar o que tem correspondência nas duas tabelas. E aí a gente pega o inner join no município status, on... No quê? Usando o quê? Usando a coluna de município ID. Então, município ID da tabela de municípios brasileiros e o município ID de município status. Beleza. Só que a gente tem uma condicional, né? Nossa condicional é que pegue, selecione apenas os dados que estão lá na nossa lista de estados. Então, a gente coloca aqui: onde municípios brasileiros estado, ou seja, onde o estado da tabela de municípios brasileiros IN (esteja em), e aqui onde está esse tanto de ponto de interrogação, quando a gente usa o format lá bonitinho, cada interrogação é substituída por um estado, né?

00:00:04:57 - 00:00:06:27

No caso, a gente usou o format, na verdade, para criar essa quantidade de interrogações que sejam exatamente a quantidade de valores dentro da nossa lista de estados. Então a gente colocou que a gente quer os estados que estejam dentro dessa lista. E, por fim, a gente faz esse group by aqui pelos estados na tabela de municípios brasileiros, tá? Fica em interrogação aqui, mas quando a gente usa o read sql aqui, que a gente passa a query, a conexão, a gente passa esse parâmetro chamado params, que é a lista de estados. E aí, quando a gente passa tudo, cada interrogação é substituída por um estado, que está dentro da lista de estados. Ok? Gente, ficou na dúvida, volta um pouquinho, vê de novo, se quiser anota direitinho: ó, nessa parte está acontecendo isso, nessa outra parte está acontecendo isso, ok? Então, beleza, a gente tem esse resultado aqui e vamos colocar esse resultado numa variável. Vou chamar ela de estados renda. Aqui... E para mergear, nós precisamos ter uma coluna em comum entre as duas tabelas. Em uma tabela, nós temos a coluna estado e na outra tabela a gente tem a coluna de UF onde mora. Para facilitar, vamos renomear a coluna de UF onde mora para chamar estado, que fica mais fácil. Então, a gente coloca dados ponto rename. E aí a gente coloca "columns", ou seja, a gente está renomeando o quê?

00:00:06:27 - 00:00:08:05

Columns igual, e aí a gente abre chaves (porque a gente vai colocar como que era a coluna antes e como que a gente quer que seja depois). Então antes estava o UF onde mora, dois pontos, e a gente quer que chame estado, ok? Aí eu vou usar o emplace igual a true, que eu basicamente estou querendo dizer para atualizar a tabela, para realmente aplicar esse rename, beleza? E aí a gente pode rodar. E quando a gente colocar dados ponto columns, a gente vai ter agora... Opa, eu coloquei errado, por isso que não teve nada. Vou copiar daqui exatamente como está o nome da coluna e colocar aqui. Agora sim, executar, executar embaixo. E agora sim, a gente tem estado. No lugar do UF onde mora, a gente tem estado. Beleza? E agora a gente pode mergear. Vamos colocar lá: dados ponto merge. A gente quer mergear com o quê? Com estados renda. Estados renda. Opa. E aí, no quê que a gente vai querer mergear? On (através da coluna de estado) ok. Como? O how. A gente vai ficar com a tabela da esquerda, que é a de dados: left. E aí a gente pode executar.

00:00:08:05 - 00:00:09:38

E se a gente for lá para o final, lá para o finalzinho da nossa tabela, a gente vê que a gente já tem uma coluna aqui que é a média lá do índice de renda por estado. Ok? Vamos atribuir esse resultado para os nossos dados mesmo. E ok. Agora a gente vai calcular a correlação de renda e salário. Em aulas passadas, nós já aprendemos a fazer o cálculo de correlação de variáveis contínuas, né? Qualquer coisa, vocês voltem na aula e assistam de novo. Nós calculamos a correlação usando o ponto cor, que é uma função do Pandas. Então, vamos lá: correlação (criando uma variável) renda e salário. E aí a gente coloca o pd ponto cor. E dentro dos parênteses a gente coloca, na verdade a gente coloca qual que é a coluna, não é pd, gente, é a coluna: dados salário ponto cor, ou seja, essa coluna eu quero fazer a correlação com qual outra coluna? Dados, e qual que é o nome da coluna de renda? Vou pegar aqui o nome das colunas de novo. É a última. Aqui, ó: AVG município renda com essa daqui. E aí a gente cola aqui dentro. Perfeito. Vou apagar essa célula de cima só pra não ficar demais. Então tá. Correlação de salário e renda.

00:00:09:38 - 00:00:10:37

E aí a gente executa. Vou dar um Enter e colocar aqui correlação renda-salário só pra poder mostrar o resultado. Então temos uma correlação aqui de 0.12, é uma correlação muito baixa entre salário e índice de renda. Em outras aulas nós já conversamos sobre o contexto, mas vale a pena pensar de novo que sabemos que existe sim uma relação entre salário e índice de renda, mas no contexto dessa pesquisa, nessa amostra de pessoas que responderam o formulário, a correlação foi quase zero. Beleza? E foi isso, pessoal. Nós finalizamos aqui mais um módulo do curso. Aprendemos o que é um banco de dados, o que é o SQL, como criar tabelas, inserir valores, fazer consultas e a conectar com o banco através do Pandas. Muita coisa, né? Mas vale lembrar de reassistir às aulas em caso de dúvidas e sempre olhar o material complementar que vai ter muita coisa legal para vocês aprenderem. Até a próxima.