

Seleccionando Campos da Base

Transcrição

[00:00] Vamos continuar o nosso projeto de análise, então agora, o que vamos fazer é uma prática bastante comum ao longo do projeto, é a seleção de algumas colunas da base de dados, para a gente não ter que ficar trabalhando a base inteira. É muito comum isso em bases muito grandes.

[00:20] Então, invés de você ficar trabalhando a base inteira toda a hora, fazendo seleções, filtros, a gente seleciona alguns campos de interesse e trabalha só aqueles campos de interesse e salva numa outra base. Essa base fica menor, fica mais rápida, então é até bastante comum essa prática de seleção de campos.

[00:40] Então, vamos fazer essa seleção aqui de campos, vou escrever aqui: “Fazendo a seleção de campos da base de dados”, só lembrando aqui, a nossa base, ela já está carregada, a gente tem 31.800, vamos trabalhar com essa base que a gente fez o tratamento da duplicidade, tiramos as duplicidades.

[01:09] A gente tem nove colunas, nove campos aqui na base de dados. Então, vamos fazer uma seleção, seleccionando alguns campos de interesse. A gente recebeu lá da área de planejamento o desafio de fazer algumas análises, justamente de procurando imóveis disponíveis.

[01:28] Então, a gente vai selecionar aqui o tipo de imóvel, então, eu vou colocar aqui, salvar numa outra nova base: “selec1” aqui, que eu vou chamar, seleção um, ela vai receber o campo... o comando que a gente utiliza para fazer essa seleção é o select.

[01:50] Então, a gente tem que colocar o nome da base, que é “aluguel_t”, vamos trabalhar com essa... “aluguel_t”, que é aluguel tratado, aonde a gente tirou as duplicidades e a gente vai e seleciona os campos da base agora. Então, a gente vai selecionar tipo.

[02:07] Só lembrando, que o R é (sensitive), então, a gente sempre tem que colocar exatamente minúscula, maiúscula para fazer essa diferenciação, senão ele não reconhece. Bairro, eu vou pegar aqui o bairro, vou colocar número de quartos, a área e o valor. Então, eu vou executar, ele vai guardar nessa base chamada select agora.

[02:47] Então, se a gente... selec1, vou dar o “Ctrl + Enter” em cima dela, seleccionou só essa base aqui ou então, melhor ainda, vamos usar o glimpse, que é para visualizar a base. Então, só para a gente ver a seleção que a gente fez aqui. Então, agora eu tenho essas cinco variáveis, mesmo volume.

[03:19] Então, quando a gente utiliza o select, a gente não está filtrando nada, a gente está apenas seleccionando o campo. Então, selecionei cinco campos: tipo, bairro, quartos, área e valor. Eu vou trabalhar com essa base aqui. Um outro tipo de seleção... um outro tipo de comando bastante comum também... outra seleção, vamos fazer uma outra seleção aqui.

[03:50] Eu poderia fazer a seleção ao contrário, eu poderia retirar da minha base alguns campos de interesse. Invés de selecionar, quando você tem muito mais campos para selecionar do que tirar, eu poderia fazer o inverso também, poderia retirar alguns campos.

[04:08] Então, vamos colocar aqui: “retirando campos da base de dados”. Para retirar os campos da base de dados ao invés de seleciona-los, eu estou fazendo uma seleção, mas eu posso retirar também, como se eu tivesse dando um drop, dropando as variáveis ou deletando alguns campos.

[04:36] Então, se... para retirar, o que que a gente faz? Então, eu vou salvar aqui de novo uma outra seleção, “selec2”, que a gente usa o mesmo comando select, aponta para a base “aluguel_t”, aluguel tratada e é só colocar um sinal de menos aqui.

[04:59] Então, se eu quiser tirar o campo IPTU, por exemplo, se eu quiser tirar ou... posso não tirar a base do IPTU, então... na sequência aqui, eu já vou dar... vou colocar... olhar a estrutura da base aqui: glimpse, essa base selec2, que a gente acabou de criar. Então, o que que aconteceu?

[05:26] Eu tinha nove campos... A minha base original eram nove campos, eu tirei da base original apenas o IPTU, então eu vou ficar com 8 campos, tipo, tipo na selec2, eu vou ter: tipo, bairro, quartos, vagas, suíte, área, valor e condomínio. Eu retirei da minha base o IPTU.

[05:48] Então, essa prática também é bastante comum, a gente pode tanto usar o select para selecionar os campos de interesse ou então, também para tirar um outro campo da minha base. Então, tem as duas maneiras. Uma outra terceira, uma outra forma de seleção, que eu vou mostrar aqui também, usando alguns outros comandos alternativos e bastante comum também.

[06:14] Então, usando outras seleções também: “usando seleção contendo alguns caracteres”. Por exemplo, imagine que você está numa base também de dados muito grande e você não lembra o nome dos campos. Então, eu poderia fazer essa seleção trazendo alguns caracteres, ele me retornando à minha base todos os campos que comesçassem ou que contenham uma determinada letra.

[06:55] Por exemplo, então a gente pode usar... eu vou... um outro exemplo aqui, select, eu vou chamar de “select3”, a mesma... eu utilizo de novo o select, comando select, aponto para a minha base: “aluguel_t”. Agora, eu uso, por exemplo, um comando chamado contains.

[07:20] Então, eu esqueci o nome do meu campo, esqueci o nome da variável. Eu vou trazer todo o... vou criar uma nova base aqui, trazendo todos os campos que contenham o campo chamado “Cond”, imagine que eu não sei o... colocando entre aspas aqui... colocar todos os campos que não lembro... que aquele campo chama condomínio, por exemplo.

[07:45] “Ah, será que tem alguma base que se chama Cond? Então, eu vou rodar isso aqui, ele rodou. E aí, se eu der o select3 aqui, vou colocar aqui o glimpse(select3), vou dar o “Ctrl + Enter”. Então, o que que ele fez? Olha, ele trouxe apenas uma coluna contendo a expressão Cond de condomínio.

[08:21] Então, isso é bastante comum também, quando está... de novo, numa base que você quer trazer alguns campos com determinadas frequências de nomes, então você pode selecionar também por uma abreviação, por números, então é bastante comum. Então, a parte de seleção era exatamente isso. Até mais.