

Agrupando uma variável condicional

Transcrição

[00:00] E se a gente quisesse filtrar nossa base para manter nela só variáveis em que as operações são válidas? No data step vimos que where não servia para filtrar a parte de variáveis que construíamos, que não estivessem presentes na base. Mas vimos no SQL que não usamos o if, usamos case when.

[00:40] Para construir o filtro, usamos o having. Tendo a flag_valida igual a um, estamos falando para manter só observações em que a flag_valida é igual a um. Antes tínhamos mais de trezentas observações, agora temos 176.

[01:17] Mas o que nós queremos é agrupar a base da mesma forma como tínhamos feito antes, com group by CPF. Tínhamos antes a variável CPF, com um em cada linha, outra variável é contagem de linhas, chamada de total_contratos. A variável de dias é construída a partir de uma operação. Podemos colocar a própria operação para fazer comparação. Se a diferença entre retorno e entrada for maior que trinta, houve um atraso e ele faz a operação.

[03:00] A variável flag_valida também precisa ter uma função de consolidação. Vamos fazer um agrupamento por CPF, e essa variável precisa ter uma função para eu consolidar. Tínhamos visto que podemos fazer uma contagem e ter a quantidade de vezes em que esse CPF aparece, podemos pegar o total da nossa flag_valida, a soma de operações que foram válidos para a nossa promoção.

[04:06] Poderíamos, pulando o passo intermediário de fazer a flag e depois fazer outro SQL com tudo, vamos fazer a sumarização já em cima do condicional. Antes dele, qual a operação de consolidação? Sum. Iremos fazer a soma da variável de condicional. Fechamos o parênteses depois do end. Estamos criando uma variável condicional, que é zero ou um dependendo do atraso e do custo de reparo, e a variável de soma desse condicional não é mais uma flag, é um total de contratos válidos.

[05:20] Temos nossa base pronta, com uma linha por CPF, mostrando que operações foram válidas. Voltando para o código, essa base que fizemos nada mais é do que para colocar na outra base de sumarização. É a mesma sumarização com variáveis a mais.

[06:02] Precisamos novamente fazer esse cruzamento, mas partindo ao invés do código que só construíamos, vamos repetir embaixo, porque vamos trocar o código que fazíamos pelo novo, que além de fazer o agrupamento pelo CPF, temos também a variável dos vários contratos válidos, que fizemos com o condicional. Iremos fazer o left join com uma variável que é o agrupamento por CPF, mas que tem também o total de contratos e os contratos válidos.

[07:58] Ele não funcionou falando que a variável não foi encontrada na base, porque nós mudamos o nome. Estamos criando uma variável que se chama total contratos e contratos válidos. Precisamos mudar também ali em cima. Vamos acrescentar à nossa base de cadastro de clientes o total de contratos, que vem da base B e uma outra variável que vem também da base B que é contratos válidos.

[08:57] Temos uma base mais robusta para essa promoção que a Alura Play quer construir. Falta comentar e salvar.