

04

Separando bases

Transcrição

[00:00] O que tínhamos feito anteriormente era gerado a base como resultado consolidado do quarto trimestre de 2017. Nisso terminamos nossa análise do quarto trimestre como um todo, somando todos os meses. O que precisamos fazer agora é analisar a quantidade de jogos dos clientes em cada um dos meses do quarto trimestre de 2017.

[01:17] Anteriormente, pegávamos o mês, summarizávamos e cruzávamos com a base de clientes. Agora, iremos fazer basicamente a mesma coisa para outubro, novembro e dezembro. Teríamos duas opções, poderíamos pegar cada um dos meses, agrupar e cruzar, ou, já que temos a base empilhada, podemos agrupar a base e depois separar de forma que tenhamos a base de cada mês já summarizada.

[02:15] Para summarizar uma base, usávamos o PROC SQL, create table, depois o nome da base, as, select, depois as variáveis que queremos ter na nossa base. Sabemos que precisávamos do CPF, mas se summarizarmos a base pelo CPF, como ela já tem diversos meses juntos, vamos ter quantos jogos o cliente alugou nos três meses de uma vez. Para distinguir, temos que summarizar também por outra variável. Sabemos que existe uma chamada safra, que diz de onde a linha saiu.

[03:33] Se summarizarmos pela safra e pelo CPF, vamos ter no final para cada uma dessas safras quantos jogos o CPF alugou. Vamos ter distinta a informação. O resultado seria a mesma coisa se tivéssemos summarizado cada base individualmente e depois empilhado. Mas agora também precisamos gerar as outras duas variáveis que tínhamos anteriormente.

[04:11] Tínhamos a contagem das linhas, como o total de contratos, a soma daquele condicional para distinguir o que é um contrato válido ou não. Isso vamos partir da nossa base de operações de 2017 quarto trimestre, que já é nossa base consolidada, e vamos agrupar pela safra e pelo CPF. Encerramos o SQL com o quit.

[05:00] Estamos vendo que cada CPF está associado a uma safra, e temos as outras duas variáveis, de total de contratos e contratos válidos. Temos o consolidado de cada CPF por safra já pronta. Só precisamos separar no código. Se for marcado por uma safra de outubro, joga para outubro. Se for novembro, joga para novembro.

[05:46] Vamos usar o data step novamente. É uma forma bastante similar ao que tínhamos feito para empilhar. Para empilhar dizíamos que queríamos criar uma base definida por três. Aqui vamos fazer o oposto. Começamos com as bases que queremos criar. Isso vamos construir a partir da base do terceiro trimestre consolidada, e precisamos falar para o SAS como ele tem que dividir. É a mesma estrutura de condicionais que víhamos utilizando. Se essa linha estiver marcada com a safra de outubro, coloque na base de outubro. Senão, verifique se a safra é de novembro, e então coloque na outra base, de novembro. Caso contrário, se a safra é de dezembro, coloque na base de dezembro. Fecho com ponto e vírgula.

[09:33] Temos uma estrutura de condicional. Encerramos o data step com run. Executando o código, temos na janela de opções as bases que criamos nessa operação.

[10:44] As bases estão consolidadas. Vamos então continuar fazendo o cruzamento de cada safra com a base de clientes, usando o create table, o nome da base, as, select todas as variáveis de A e as variáveis de B. Mas vamos renomear, porque não é mais do quarto trimestre. É o total de contratos de outubro. Iremos partir cruzando com a base de cadastro de cliente v3, que está no diretório da Alura e vai ser chamada de A. Vamos fazer um cruzamento à esquerda com a base que tínhamos gerado antes.

[12:52] Quais são nossas chaves de cruzamento? É o CPF da base de clientes transformado em numérico e usamos isso para cruzar com a base de operações. Terminamos com o quit.

[14:00] Gerei minha base de cadastro de clientes jogos de 201710, que tem o total de contratos e total de contratos válidos. Fizemos isso para uma das safras. Precisamos repetir para as demais.