

10

Juntando informações com o cruzamento de bases

A AluraPlay deseja fazer uma campanha com os clientes que depende da quantidade de jogos que cada um deles alugou. As informações dos aluguéis de jogos em si não está em nenhuma das bases que trabalhamos até agora, mas sim nas bases de operações, que são bases geradas mensalmente e que contém as informações sobre os contratos de aluguel de jogos em si. Vamos trabalhar aqui com a base de operações de setembro de 2017, chamada `OPERACOES_201709` .

Dado que essas bases de operações possuem uma variável que contém o CPF do cliente que alugou um determinado jogo, e que cada linha dessa base é um jogo alugado, se summarizarmos ou agruparmos essa base pela variável CPF e criarmos uma variável que contém a quantidade de linhas por CPF, teremos a informação de quantos jogos cada cliente alugou (pois cada CPF representa um cliente).

Já vimos como agrupar bases usando, por exemplo, o `PROC SQL` . O código abaixo faz esse agrupamento que precisamos:

```
PROC SQL;
  create table contratos_cpf as
    select CPF, count(*) as Quantidade_jogos
    from alura.operacoes_201709
    group by 1
;QUIT;
```

Mas nessa base gerada há apenas uma informação do cliente: o CPF, e isso não é suficiente para a área de relacionamento com os clientes da empresa, eles precisam que esta informação de quantidade de jogos que cada cliente alugou seja adicionada à base de cadastro dos clientes, pois é nela que estão outras informações muito importantes, como nome e forma de contato. Precisamos juntar essas informações em uma base única.

Esse processo de juntar informações (variáveis) de bases distintas em uma única base é chamado, muitas vezes, de **cruzamento** de bases, onde fazemos a associação de uma chave comum entre as bases para que seja possível uma correspondência entre as observações de cada base.

Usando um *data step*, faça o cruzamento da base `CONTRATOS_CPF` , gerada pelo código acima, com a base `CADASTRO_CLIENTE_V3` , pela chave `CPF` .

Obs: A forma que o CPF está armazenado em cada base é diferente: Na base de clientes temos o CPF completo em formato caractere; na base de operações temos o CPF raiz (sem os últimos dois dígitos verificadores) em formato numérico. Antes de fazer o cruzamento, estas chaves precisam ser trabalhadas para que fiquem em formatos similares.