

Checagem de bases

Transcrição

[00:00] Vimos que nosso resultado está aparecendo pela saída da nossa listagem das bases, mas o cliente tinha pedido um pouco a mais de informações, ele queria que a gente checasse a integridade da base para ver se ela realmente tem os clientes que ele sabe que tem no cadastro, se a base tem todos os produtos. Precisaríamos checar a quantidade de linhas que essas minhas bases têm.

[00:35] O dataset que eu usei não está me dando essas informações. Lembrando que o PROCDATASET é o processo para fazer a listagem das bases que eu tenho no diretório. Dessa forma mais crua ele não está me dando todas as informações, quero mais detalhes. Posso pedir isso se eu colocar a opção dentro do código. Posso passar meu parâmetro de declarar qual a biblioteca que quero que ele me liste e acrescento a opção de details.

[01:41] Já mudou minha saída. Podem perceber que tem uma coluna de observações e uma de variáveis. Ele me diz que cadastro de cliente tem cem observações e oito variáveis e cadastro de produto tem quinhentas observações e seis variáveis.

[02:05] Vimos que o dataset estava chamando de observações e variáveis, mas o que são elas? Quando estamos no dia a dia, temos a lista de contatos no celular, com várias pessoas que conhecemos e informações sobre elas. Por exemplo, posso ter o nome da pessoa, o telefone e o e-mail. Sem saber, essa lista de contatos já é uma base de dados, uma forma de organizar informações.

[03:14] Se lembrarmos um pouco da matemática, vemos uma estrutura chamada matriz, que é a mesma coisa que nossa planilha, nossa lista de contatos. Você organiza as informações em linhas e colunas. Ou seja, minha matriz também é uma base de dados, se eu relembrar a estrutura da lista de contatos. Eu sei que linha 1 e coluna 1 tem o nome de fulano, linha 1 e coluna 2 é o telefone dele, e assim por diante.

[04:00] Um estatístico também faz uso de organizar as informações dessa maneira. Mas da mesma forma que vemos nomes diferentes para uma mesma estrutura, o estatístico também faz isso. No caso, o que chamávamos de matriz, ele vai chamar de tabela. O que chamávamos de linha, ele vai chamar de observação, e as colunas de variáveis. Continua sendo a mesma coisa. Minha tabela é uma matriz, que é uma base de dados.

[04:55] Voltando para o SAS, já entendi o que as informações significam. Quando ele diz que a base de cadastro de clientes tem cem observações, ele me diz que ela tem cem linhas. Oito variáveis, oito colunas.

[05:20] Voltando para a aba do código, preciso conseguir informações a mais sobre a base em si. Já chequei que ela está com o número de linhas e colunas que o cliente tinha falado, mas não sei se estão corretas. Preciso de outro comando para ver o conteúdo de uma base específica. É um comando que eu sei que é corriqueiro, muita gente usa, vai ser um procedimento também pronto do SAS.

[06:10] Começo escrevendo PROCCONTENTS. Ele vai me dar o conteúdo de uma base específica. Não vou usar o mesmo comando de lib, porque quero informação de uma base. Em inglês, a base normalmente é chamada de data ou dataset, mas o SAS resume para data. Vai ser igual ao nome da minha base cadastro produto. Encerrado o comando, coloco ponto e vírgula e executo.

[07:20] Ele me dá um erro e diz que a base não existe. Isso é estranho, acabei de ver que existe. Mas existe onde? Lembra que no começo tínhamos feito a declaração em que queria associar o diretório físico ao nome Alura? Para especificar o

diretório físico em que minha base está não preciso mais ficar escrevendo o caminho físico. Eu já dei o atalho. Na hora de chamar a base, só preciso dizer para o SAS antes em que diretório ele está.

[08:14] Uma coisa é o diretório, outra é a base que está lá dentro. Como falo que quero a base de cadastro de produto que está dentro do diretório Alura? Depois da biblioteca, coloco um ponto. Isso quer dizer que dentro desse diretório do Alura, quero a base de cadastro de produto. Ou seja, diretório ponto a base que está dentro daquele diretório.

[08:52] Vou selecionar novamente, pedir para executar, e agora vai corretamente. Já tenho uma saída em que no canto superior esquerdo tenho observações e variáveis, e para baixo tenho a tabela com a lista de variáveis, tipo de variável, tamanho, formato e label.

[09:28] Tenho as variáveis chamadas data, gênero, nome, número, plataforma e preço. Tanto número quanto preço estão sem cê-cedilha e sem acentuação, porque no SAS não podemos ter caracteres especiais. Mas a pessoa que criou a base para nós já ficou preocupada em criar um rótulo explicando mais sobre a variável. Data é a data de lançamento, número é o código do produto, preço é o preço.

[10:28] Já sei que essas são minhas variáveis. Sei o que elas significam. Quero ver minha base em si. Já sei que tenho uma coluna com o nome do jogo, mas quero ver os nomes dos jogos. Posso usar um comando para imprimir a base. Vai ser o PROCPRINT. Meu parâmetro novamente é uma base específica. Uso o comando de data, dizendo onde está a base, e qual das bases do diretório quero. Encerro com ponto e vírgula, depois run e ponto e vírgula.

[11:56] Tenho minha base de cadastro produtos. Tenho a variável número, o nome do jogo, o gênero, a plataforma, a data de lançamento e o preço.