# Dicionário de dados

Um dicionário de dado armazena e comunica informações sobre um banco de dados. É considerado uma lista de dados com os principais termos e métricas do banco de dados que está vinculado. Esse material também pode ser entendido como um glossário dos seus dados. Embora pareça simples, ele é crucial para alinhar o funcionamento da equipe que irá trabalhar com os dados e manter todas as informações padronizadas. O conteúdo do dicionário de dados pode variar, mas normalmente inclui alguns itens como:

* Nomes e definições das variáveis;
* Propriedades detalhadas dos dados (tipos de dados, tamanho, se há valores nulos);
* Sequência das perguntas ou da coleta das variáveis;
* Dados de referência (para fazer classificações e domínios descritivos);
* Codificação para dados ausentes.

## Variável

Um dos primeiros passos na confecção de uma pesquisa ou levantamento de dados é definir o que será avaliado, ou seja, quais são as variáveis que irão ajudar a responder a minha pergunta principal. Ao determinar quais são as variáveis que irão compor o seu banco de dados, você já pode iniciar a confecção do seu dicionário de variáveis.

## Nome da Variável

O nome da variável representa o rótulo\* (ou label) que você quer atribuir a sua variável e que representará a variável no banco de dados. Ao contrário do item anterior, nesse item você deve colocar um nome simples, sem acentuação, caracteres especiais e sem espaços entre palavras.

Por exemplo, imagine que você irá utilizar uma escala psicométrica e cada item dessa escala (E) será uma variável. No nome da variável você pode colocar um código simples que presente a variável, como: E1 e E2.

## Tipo de Variável

Saber o tipo de variável que contém no banco é fundamental para analisar os dados. Muitas vezes é necessário recalcular uma variável para deixá-la no formato ideal, esse cenário poderia ser evitado com um bom planejamento da coleta de dados e a utilização de dicionários de dados.

Uma variável pode ser numérica ou categórica. Em especial nas variáveis categóricas, é importante especificar as categorias e se essas categorias possuem uma ordem de importância ou não. Então, sinalize sempre no tipo de variável se ela é numérica ou categórica, especifique ainda se as categorias possuem uma ordem de importância ou hierárquica (como, escolaridade ou grau de uma doença). Também há variáveis do tipo tempo (ou time), que representa alguma informação de tempo, como: segundos, minutos, horas, datas. Nesses casos, seria importante discriminar qual das medidas de tempo estão representadas na variável.

## Descrição

O espaço para descrição é destinado para incluir maiores detalhes sobre determinada variável. Uma descrição detalhada é encorajada para que todos possam entender o que realmente aquela variável representa, qual a razão dela está ali e como deve ser tratada durante as análises.

Você pode acrescentar também detalhes de como e quando essa variável foi coletada ou em que momentos ela não será coletada. Ao mesmo tempo, evite descrições desnecessárias e circulares, como por exemplo: “a idade do paciente representa quanto tempo o paciente está vivo”. Diga informações diretas e relevantes, como: “a idade fornecida pelo paciente em anos”. Dessa forma, você sabe que a ideia foi fornecida diretamente os dados da variável em questão foi fornecido pelo paciente e está em anos.

## Valores Permitidos

No campo de valores permitidos você deve fornecer todos os valores que são possíveis de serem coletados. Para as variáveis numéricas indique o valor mínimo e máximo que podem ser informados no seu banco. Já para as variáveis categóricas informe o número da categoria, bem como o rótulo associada a ela.

Exemplos: 2 categorias: X e Y, 001-999, 27 estados, até 30 meses e ID (Numero incremental)

## Possui Valores Nulos

Durante o processo de análise de dados os valores faltantes sempre são um problema. Há momentos em que não é possível coletar determinada variável e os dados daquele caso ficarão em branco ou vazio. Isso é perfeitamente possível no mundo real, mas é um problema durante o processo de análise de dados. Portanto, ao fazer o exercício da construção de um dicionário de variáveis você pode evitar dados nulos. Mas se não há saída, sinalize no seu dicionário de dados que aquela variável pode ter valores nulos ou faltantes.

## Anotações

Ao final do seu dicionário de dados você pode criar um campo de anotações. Coloque nesse campo demais informações que não foram até então declaradas. Como, por exemplo, a referência que você utilizou para criar as categorias de uma variável ou o estudo que tem uma variável parecida e que fomenta a discussão dos seus resultados.

Exemplos: O item foi o resultado de uma pesquisa X, a variável constante com relação ao tempo ou é uma resposta de algum pré-evento ou pós-evento (Evento: Cadastro no sistema ou Realização de algum pagamento).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variável | Nome da variável | Descrição | Tipo de dado | Valores permitidos | Possui valores nulos? | Anotações |
| Exemplo:  Número do CPF | Exemplo:  CPF | Exemplo:  Identificação exclusiva do cliente | Exemplo: Inteiro | Exemplo:  001-999 | Exemplo:  Não |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |