

# Tarefa 1 – Base de Dados

Peterson Junior

3 de setembro de 2025

## 1 Introdução

Nesta primeira tarefa, o objetivo é escolher uma base de dados de interesse, realizar sua importação no R e conduzir uma análise exploratória inicial. Com isso, busca-se compreender a estrutura da base, suas variáveis e preparar um *dicionário de dados*, ferramenta essencial para qualquer trabalho científico que envolva dados.

## 2 Banco de Dados

A base escolhida foi **Health\_Risk\_Dataset.csv**, que contém informações de pacientes com diferentes sinais vitais, estado de consciência e nível de risco associado. A escolha se justifica pelo fato de ser uma base que permite análises relevantes em saúde, especialmente no contexto de monitoramento clínico e avaliação de risco hospitalar. Essa base é suficientemente rica em variáveis numéricas e categóricas, além de incluir uma variável alvo (**Risk\_Level**) que pode ser utilizada em modelos de aprendizado de máquina para classificação.

### Procedimentos de Importação no R

O carregamento dos dados foi feito utilizando o pacote **readr**, conforme instruções do enunciado:

```
library(readr)
health <- read_csv("Health_Risk_Dataset.csv")
head(health, 10) # primeiras linhas
tail(health, 10) # últimas linhas
str(health)      # estrutura da base
summary(health)  # resumo estatístico
```

## 3 Variáveis

A base contém 1000 registros e 10 variáveis. A seguir, apresenta-se o dicionário de dados com informações sobre cada variável, seu tipo, unidade, descrição e observações importantes.

Nome da Variável	Tipo	Unidade	Descrição	Observação
Patient ID	Categórica (texto)	Não se aplica	Identificação única de cada paciente.	Serve como chave, não deve ser usada em análises estatísticas.
Respiratory Rate	Numérica (inteiro)	Respirações/min	Taxa respiratória do paciente.	Valores muito altos ou baixos indicam risco.
Oxygen Saturation	Numérica (inteiro)	%	Saturação de oxigênio no sangue.	Normal $\geq 95\%$ , valores baixos sugerem hipóxia.
O2 Scale	Numérica (inteiro)	Escala (1–3)	Escala de suporte de oxigênio utilizado.	Usado para classificar necessidade de oxigênio suplementar.
Systolic BP	Numérica (inteiro)	mmHg	Pressão arterial sistólica.	Valores muito baixos indicam choque; muito altos, hipertensão.
Heart Rate	Numérica (inteiro)	bpm	Frequência cardíaca.	Frequência elevada pode indicar estresse fisiológico.
Temperature	Numérica (decimal)	°C	Temperatura corporal.	Febre ( $> 38^{\circ}\text{C}$ ) ou hipotermia ( $< 36^{\circ}\text{C}$ ) são sinais críticos.
Consciousness	Categórica (texto)	Não se aplica	Estado de consciência do paciente (A = Alerta, V = Verbal, P = Dor, U = Inconsciente).	Escala simplificada de Glasgow.
On Oxygen	Numérica (binária)	0 = Não, 1 = Sim	Indica se o paciente está em uso de oxigênio suplementar.	Binária, pode ser tratada como fator no R.
Risk Level	Categórica (texto)	Não se aplica	Classificação de risco do paciente (Low, Medium, High).	Variável-alvo para estudos de classificação.

## 4 Conclusão

O estudo inicial da base de dados permitiu compreender a estrutura das variáveis e organizar um dicionário detalhado. Esse material será essencial para futuras etapas de análise estatística e modelagem, especialmente em problemas de classificação de risco em saúde.