Explicação sobre alteração da tabela

No início do trabalho, carregamos os dados diretamente de um arquivo CSV para dentro de um DataFrame do **Pandas**, que é como uma planilha em memória. Isso foi feito com:

import pandas as pd

df = pd.read\_csv("dados.csv")

A partir daí, começamos a limpeza dos dados. A primeira etapa foi tratar os valores ausentes. Muitos modelos de machine learning e até mesmo operações simples do Pandas não funcionam bem quando há NaN. Para evitar erros, substituímos valores nulos por zero:

df = df.fillna(0)

Depois, trabalhamos com a **coluna de datas**. Como os dados vinham em formato de texto (string), era preciso transformar em datas reais para facilitar cálculos ou filtros por período. Fizemos a conversão usando:

df['data'] = pd.to\_datetime(df['data'], errors='coerce')

O parâmetro errors='coerce' garantiu que qualquer valor inválido fosse convertido para NaT (um tipo de nulo para datas), em vez de gerar erro.

Na sequência, tratamos a **coluna de idades**, que também vinha com problemas de formatação. Para isso, transformamos tudo em números inteiros e eliminamos valores inválidos:

df['idade'] = pd.to\_numeric(df['idade'], errors='coerce')

df = df.dropna(subset=['idade'])

Assim, só ficaram idades válidas e numéricas, o que garante consistência nas análises.

Outro ponto importante foi a coluna **“ativo”**, que estava com valores escritos de várias formas diferentes, como "Sim", "sim", "S", "Não", "nao", etc. Padronizamos tudo para **booleano** (True ou False), com o seguinte código:

df['ativo'] = df['ativo'].str.strip().str.lower()

df['ativo'] = df['ativo'].map({'sim': True, 's': True, 'não': False, 'nao': False, 'n': False})

Aqui usamos .str.strip().str.lower() para tirar espaços e deixar todas as letras minúsculas. Depois, usamos .map() para transformar os textos em valores padronizados.

Por fim, após todas essas alterações, os dados ficaram mais **limpos, padronizados e prontos para uso**, seja para análises estatísticas, relatórios ou até como base para treinar modelos de machine learning.