#### **1. Carregamento dos Dados**

* O arquivo dados\_nao\_tratados.csv foi carregado usando pandas.read\_csv() para um DataFrame df.

#### **2. Conversão de Colunas para Tipo Numérico**

* As colunas nota e idade foram convertidas para valores numéricos usando pd.to\_numeric() com errors='coerce', que transforma valores inválidos em NaN.

#### **3. Preenchimento de Valores Ausentes**

* Todos os valores NaN foram substituídos por 0 com df.fillna(0, inplace=True).

**4. Normalização da Coluna ativo**

* A coluna ativo foi padronizada para valores booleanos (True ou False) com o método replace():  
    
   **Valores substituídos:**

| **Valor Original** | **Valor Convertido** |
| --- | --- |
| "Sim", "sim", "e", "Yes", "TRUE" | True |
| "Não", "nao", "FALSE" | False |

**5. Tratamento da Coluna data\_inscricao**

* Algumas entradas foram corrigidas manualmente com replace() para garantir um formato de data válido.
* Após substituição, a coluna foi convertida para o tipo datetime com o formato %Y-%m-%d e errors='coerce'.

#### **6. Correção de Valores Específicos**

* **Idade:**
  + Valores 0 foram substituídos por '21' e convertidos novamente para numérico.
* **Nota:**
  + Valores 0 foram substituídos por '7.5' e convertidos novamente para numérico.