

Documentação Técnica – Tratamento de Dados

1. Introdução

Este documento descreve as etapas de tratamento de dados realizadas no notebook 'dados_ao_tratados.ipynb'. O objetivo principal é preparar a base de dados bruta para análises, garantindo consistência, padronização e melhor qualidade dos dados. O notebook foi desenvolvido em Python, utilizando a biblioteca Pandas para a manipulação de dados. A base de dados e o carregamento do arquivo 'dados_ao_tratados.csv'.

2. Etapas de treinamento

2.1 Importação de Bibliotecas

Foi importada a biblioteca Pandas para a manipulação e transformar dados.

```
import pandas as pd
```

2.2 Carregamento do Dataset

O dataset foi carregado a partir do arquivo (.CSV)

```
df = pd.read_csv('dados_ao_tratados.csv')
df
```

2.3 Tratamento da coluna “ativos”

A coluna continha valores inconsistentes, "e", "yes", "sim", "true", "false", "não", "nao". Esses valores foram transformados para valores booleanos (True/False).

```
df["ativo"] = (
    df["ativo"]
    .str.strip()
    .str.lower()
    .map({"e":True,"yes":True,"sim":True,"true":True,"false":False,"não":False,"nao":False})
)
df
```

2.4 Conversão da coluna “idade”

A coluna 'idade' foi convertida para valores numéricos, eliminando os possíveis inconsistências ou valores inválidos.

```
df["idade"] = pd.to_numeric(df["idade"], errors="coerce")
df
```

2.5 Conversão da Coluna “data_incritos”

A coluna ‘data_incritos’ foi convertida para o formato datetime, considerando o padrão de datas brasileiro dia/mês/ano.

```
df["data_inscricao"] = pd.to_datetime(df["data_inscricao"], errors="coerce", dayfirst=True)
```

3. Conclusão

Após esses tratamentos, o dataset está padronizado, pronto para ser utilizados em análise estatísticas. As principais tratamentos desse dataset foram padronização de variáveis categóricas, conversão de tipos de dados e correção de inconsistências