



# Bootcamp: Data Science

## Chat de Debates

Módulo: Análise Exploratória de Dados

### Enunciado do Chat de Debates

Oi, turma! Prof. Pedro Calais aqui.

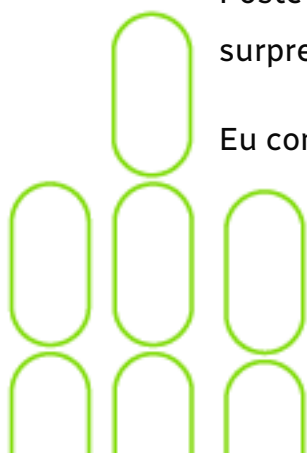
Espero que o módulo sobre como conduzir análises exploratórias com dados esteja sendo útil para você.

Pensei em algo para interagirmos; escolhi um conjunto de dados públicos, o [Censo Escolar 2023](#), e quero lançar o seguinte desafio: cada um de vocês vai ler os dados com o Pandas, no Python, e executar alguma análise exploratória que gere alguma tabela, visualização ou estatística descritiva de seu interesse. Ao invés de programar diretamente uma análise, você pode também gerar os relatórios das ferramentas que conhecemos no módulo (como sweetviz e ydata-profiling) e trazer aqui. Não se esqueça de dar sua interpretação para análise ou explicar que tipo de pergunta de negócio esta análise pode ajudar a responder.

Poste no fórum o resultado da sua análise e comente se o resultado te surpreendeu ou se você aprendeu algo com ele.

Eu começo! Li arquivo com o *pandas* assim:

```
import pandas as pd
```



```
escolas=pd.read_csv("C:/Users/pedro/Downloads/microdados_censo_escolar_2023/microdados_censo_escolar_2023/dados/microdados_ed_basica_2023.csv", encoding='latin-1', sep=';')
```

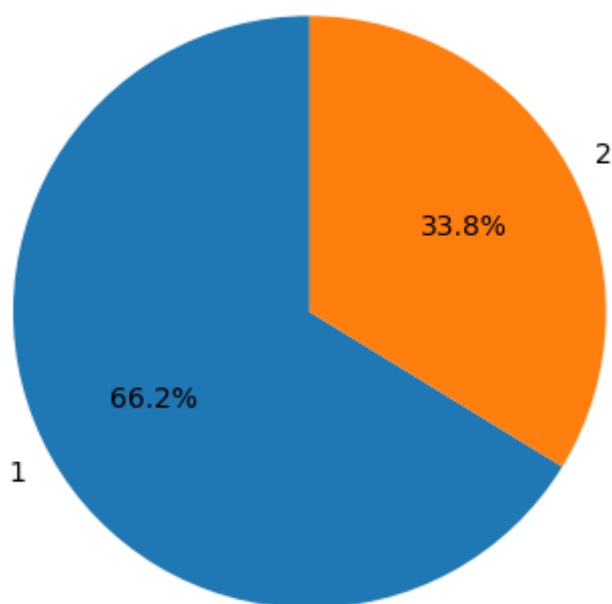
*Você pode executar este código no Google Colab ou Visual Studio Code. No caso do Google Colab, suba o arquivo para a área de arquivos do Google Colab.*

Depois, pedi para o chatGPT gerar o seguinte código para mim:

*Usando o pandas, gere um gráfico de pizza com os atributos de um campo de nome TP\_LOCALIZACAO.*

Copiei o código e, em menos de um minuto, plotei o seguinte gráfico:

Distribuição de TP\_LOCALIZACAO



Achei interessante ver que o número de escolas na zona rural representa um terço do total de escolas – um número maior do que eu esperava. Pronto! Aprendi algo sobre o panorama geral das escolas no Brasil.

O arquivo *dicionário\_dados\_educação\_básica.xlsx*, contido no arquivo de microdados do Censo Escolar 2023, que você encontra [aqui](#), explica o que é cada



um dos 458 campos da planilha – não faltam oportunidades de análise para vocês explorarem! Podem comentar as análises dos colegas e se inspirarem, mas é importante que cada um conduza uma análise diferente.

Um abraço,

Prof. Pedro.

