

[Entrar](#)[Cadastrar](#)

Esta é uma cópia de uma conversa entre ChatGPT e Anônimo.
[Denunciar conteúdo](#)

📎 Arquivo carregado

Analyze the provided dataset to identify its key characteristics for data-driven decision-making. Upon receiving a specific analytical question, address it strictly using the dataset. Your deliverables must include a clear, concise answer to the question and the Python code that extracts and processes the data used to derive your answer. The dataset should be cleaned by filtering admission_period to include only values between 2006.1 and 2019.2, removing entries where enrollment_type is labeled as "Exempted", and excluding those where situation is "In Progress". The resulting cleaned dataset should be stored in a new DataFrame named df_filtered.

I'll start by loading the dataset, inspecting its structure, and then applying the required cleaning steps. Let's begin.

The dataset contains 30,130 entries and 34 columns, including details about admission, dropout, enrollment, and academic performance. Now, I'll apply the required cleaning steps:

1. Filter admission_period: Keep only records between 2006.1 and 2019.2.
2. Remove "Exempted" enrollment_type.
3. Exclude entries where situation is "In Progress".

I'll proceed with these steps now. [↗]

The cleaned dataset df_filtered now contains 16,563 entries. Let me know your specific analytical question, and I'll analyze the data accordingly! [↗]

What is the overall average number of credit hours taken by students in their first semester, across all years in the dataset?

To determine the overall average number of credit hours taken by students in their first semester, I will:

1. Identify students' first semester by selecting records where relative_period == 1.0.
2. Calculate the average credits for these students.

I'll proceed with the analysis now.

The overall average number of credit hours taken by students in their first semester is 3.87.

Here's the Python code used for this analysis:

python

📄 Copiar

✎ Editar

```
# Filter for students in their first semester (relative_period == 1.0)
df_first_semester = df_filtered[df_filtered["relative_period"] == 1.0]

# Calculate the average number of credit hours taken in the first semester
average_credits_first_semester = df_first_semester["credits"].mean()
average_credits_first_semester
```

Let me know if you need any further insights! [↗]

Envie uma mensagem para o ChatGPT

📎 Anexar

🗣️ Voz

O ChatGPT pode cometer erros. Considere verificar informações importantes.

?