

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

Esta é uma cópia de uma conversa entre ChatGPT e André.

[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 first inspect the dataset to understand its structure and content. Then, I'll apply the specified cleaning criteria:

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

Let's start by loading and examining the dataset.

The dataset contains 30,130 entries and 34 columns. Key columns relevant for filtering are:

- admission_period (float): Needs to be between 2006.1 and 2019.2.
- enrollment_type (object): Exclude rows where it's "Exempted".
- situation (object): Exclude rows where it's "In Progress".

Now, I'll apply these filters and create df_filtered. ↴

The dataset has been successfully filtered, reducing it from 30,130 entries to 10,305.

Now that we have df_filtered, what's your specific analytical question? ↴

Calculate the dropout frequency for each academic period, and identify any noticeable differences between semesters and years.

To analyze dropout frequency across academic periods, I'll:

1. **Count the number of dropouts per dropout_period.**
2. **Compare dropout rates between different years and semesters.**

Let's proceed with the analysis.

Dropout Frequency Analysis:

1. **Yearly Trends:**

- Dropouts increased steadily from 2006 (2 dropouts) to a peak in 2019 (1,683 dropouts).
- A significant number of records are associated with year 0 (4,002 dropouts), possibly indicating missing or improperly recorded years.
- A sharp drop is observed in 2020 (232 dropouts), likely due to the COVID-19 pandemic.

2. **Semester Comparison:**

- First semester (0.1): 6,348 dropouts
- Second semester (0.2): 6,213 dropouts
- Unspecified (0.0): 4,002 dropouts
- Dropouts appear evenly distributed between the two semesters, with a slight increase in the first semester.

Would you like a visualization of this data for better insights? ↴

Envie uma mensagem para o ChatGPT

@ Anexar

Voz

© ChatGPT pode cometer erros. Considere verificar informações importantes.

?