

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

Esta é uma cópia de uma conversa entre ChatGPT e Andréino.
[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 begin by inspecting the dataset to understand its structure and contents. Then, I'll apply the specified cleaning steps:

1. Filter `admission_period` to include only values between 2006.1 and 2019.2.
2. Remove entries where `enrollment_type` is labeled as "Exempted".
3. Exclude rows where `situation` is "In Progress".

Let's start by loading and examining the dataset.

The dataset contains 30,130 entries and 34 columns. Key observations:

- `admission_period` is a float and includes values outside the specified range.
- `enrollment_type` has an "Exempted" category that needs to be removed.
- `situation` includes an "In Progress" category that must be excluded.

I'll now clean the dataset according to your specification. ↓ I'll store the result in `df_filtered`. [↩]

The cleaned dataset (`df_filtered`) contains 10,365 entries, down from the original 30,130, after applying the filtering criteria:

- `admission_period` is restricted between 2006.1 and 2019.2.
- Entries with `enrollment_type` as "Exempted" were removed.
- Entries where `situation` is "In Progress" were excluded.

The dataset is now ready for analysis. Please provide a specific analytical question you'd like me to address using this data. [↩]

Calculate the entropy of the distribution of students across academic departments for each of the last 5 recorded semesters, and describe the trend over time.

To calculate entropy for the distribution of students across academic departments over the last five recorded semesters, I'll follow these steps:

1. Identify the five most recent semesters from the `admission_period` column.
2. Group students by `sector_name` (academic department) within each semester.
3. Compute the probability distribution of students across departments for each semester.
4. Use Shannon entropy formula:

$$H = - \sum p_i \log_2(p_i)$$

where p_i is the proportion of students in each department.

5. Analyze the trend over time. ↓

Let's perform the calculations.

Here are the entropy values for the distribution of students across academic departments in the last five recorded semesters:

Admission Period	Entropy
2017.2	1.493
2018.1	1.210
2018.2	1.445
2019.1	1.434
2019.2	1.576

Trend Analysis:

- 2018.1 shows a noticeable drop in entropy (1.210), indicating a more concentrated student distribution in fewer departments.
- A general increasing trend from 2018.1 to 2019.2 suggests that student enrollment became more evenly distributed among departments.
- The highest entropy (1.576 in 2019.2) suggests that, in this semester, students were spread across a greater variety of departments more evenly than in prior semesters.

Would you like a visualization of this trend? [↩]



Envie uma mensagem para o ChatGPT

📎 Anexar

🗣️ Voz

© ChatGPT pode cometer erros. Considere verificar informações importantes.

?