



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 start by loading the dataset, examining its structure, and applying the necessary cleaning steps. Let's proceed.

The dataset contains 30,130 entries and 34 columns. The key steps for cleaning include:

1. Filtering `admission_period` to only include values between 2006.1 and 2019.2.
2. Removing rows where `enrollment_type` is labeled as "Exempted."
3. Excluding rows where `situation` is "In Progress."

Now, I'll apply these filters and create `df_filtered`. [↵]

The cleaned dataset, `df_filtered`, contains 16,563 entries and retains all 34 original columns. Let me know your specific analytical question, and I'll extract the relevant insights. [↵]

Is there a statistically significant difference in grades between students enrolled in Normal and Extracurricular modes?

To determine if there is a statistically significant difference in grades between students enrolled in **Normal** and **Extracurricular** modes, we will:

1. Extract the `final_average` (grades) for both enrollment types.
2. Check for normality to determine if we should use a parametric or non-parametric test.
3. Conduct an appropriate statistical test (e.g., **t-test** if normally distributed, **Mann-Whitney U test** otherwise).

Let's proceed with the analysis.

The **Mann-Whitney U Test** was used since the grade distributions were not normally distributed. The p-value is 2.8×10^{-7} , which is far below the common significance threshold of 0.05.

Conclusion: There is a **statistically significant** difference in grades between students enrolled in **Normal** and **Extracurricular** modes. [↵]

Envie uma mensagem para o ChatGPT

@ Anexar

🔊 Voz

© ChatGPT pode cometer erros. Considere verificar informações importantes.

?