



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 and inspecting the dataset to understand its structure and contents. Then, I'll apply the specified cleaning steps before storing the filtered data in `df_filtered`. Let me proceed with this now.

The dataset contains 30,130 entries with 34 columns. Key columns relevant to our cleaning process include:

- `admission_period` (float64): We need to filter this between 2006.1 and 2019.2.
- `enrollment_type` (object): We will exclude rows labeled "Exempted."
- `situation` (object): Rows labeled "In Progress" will be removed.

Now, I'll clean the dataset and store the result in `df_filtered`. [↩]

The dataset has been successfully filtered, reducing the number of entries from 30,130 to 16,563. It now includes only records where:

- `admission_period` is between 2006.1 and 2019.2.
- `enrollment_type` is not "Exempted."
- `situation` is not "In Progress."

You can now provide a specific analytical question, and I'll extract insights using this cleaned dataset. [↩]

Analyze how a student's age at entry and dropout influences their overall average (mean of non-null grades). Apply one statistical method (e.g., regression or correlation) to measure significance, considering confounding factors.

To analyze the influence of a student's **admission age** and **dropout age** on their **overall average** (mean of non-null grades), I will:

1. Compute each student's mean of non-null `final_average` scores.
2. Perform a correlation analysis and a multiple linear regression to measure the statistical relationship.
3. Interpret the significance of the findings while considering potential confounding factors.

I'll proceed with the analysis now.

Analysis Results

1. Correlation Analysis

- **Admission Age vs. Final Average**
 - Correlation: -0.0136 (very weak negative)
 - p-value: 0.587 (not statistically significant)
- **Dropout Age vs. Final Average**
 - Correlation: 0.2077 (moderate positive)
 - p-value: 4.57e-17 (highly significant)

2. Multiple Linear Regression

- **Model $R^2 = 0.221$** , meaning 22.1% of the variation in final average is explained by admission and dropout age.
- **Key findings from regression coefficients:**
 - Admission Age (-0.5963, $p < 0.001$) → Older students at entry tend to have lower final averages.
 - Dropout Age (+0.6487, $p < 0.001$) → Higher dropout age is associated with higher final averages.

Conclusion

- Admission age has little impact, but slightly older students at entry may perform worse.
- Dropout age significantly correlates with better final averages, likely because students who persist longer in the program gain more experience or are closer to graduation.
- Confounding factors like academic background, course difficulty, or motivation could also be influencing these results.

Would you like a visualization of these relationships? [↩]

Envie uma mensagem para o ChatGPT

@ Anexar

Voz

O ChatGPT pode cometer erros. Considere verificar informações importantes.

?