



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 and inspecting its structure. Then, I'll apply the cleaning steps as specified:

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

Let's begin:

The dataset contains 30,130 rows and 34 columns. Now, I'll apply the required filtering steps to create

df_filtered:

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

Let's proceed with these filtering steps. (-)

The filtered dataset (df_filtered) now contains 16,563 entries, ensuring that only relevant data is included.

Let me know your specific analytical question, and I'll extract insights strictly using this dataset. (-)

Analyze the relationship between extra-curricular participation (with no failures or dropouts) and academic success using correlation tests (Pearson, Spearman), regression models, and propensity score matching. Compare the effectiveness of each approach.

To analyze the relationship between extra-curricular participation and academic success, we will follow these steps:

1. Define Key Variables

- Independent Variable (Treatment): Participation in extra-curricular activities.
- Dependent Variable (Outcome): Academic success (e.g., final average, graduation status).

2. Data Preparation

- Identify students who participated in extra-curricular activities.
- Ensure they have no failures or dropouts.
- Define academic success (e.g., final average score or graduation).

3. Analysis Methods

- Correlation Tests (Pearson & Spearman): Measure the association between participation and academic success.
- Regression Models: Assess the effect of participation while controlling for confounding variables.
- Propensity Score Matching (PSM): Match students with and without participation based on similar characteristics to estimate causal impact.

Step 1: Identify Relevant Data Columns

I'll now inspect the dataset to identify columns related to extra-curricular participation, failure/dropout status, and academic success before proceeding with the analysis.

Relevant Columns for Analysis

1. Extra-curricular Participation: This dataset does not have an explicit column for extra-curricular activities. Do you have a specific indicator for this, or should we infer it based on subject names, workload, or another attribute?
2. No Failures or Dropouts:
 - situation should indicate students who graduated successfully (excluding failures or dropouts).
3. Academic Success:
 - final_average (student's final grades).
 - exit_form to check if they graduated.

Before proceeding with correlation tests, regression, and PSM, do you want me to infer extra-curricular participation based on certain subjects or another characteristic? (-)

Analyze the relationship between extra-curricular participation (with no failures or dropouts) and academic success using correlation tests (Pearson, Spearman), regression models, and propensity score matching. Compare the effectiveness of each approach.

Step 1: Identifying Extra-Curricular Participation

The dataset does not have a direct indicator for extra-curricular activities. To proceed, I will infer extra-curricular participation based on subject names that suggest involvement in activities beyond standard coursework, such as:

- Internships (INTEGRATED INTERNSHIP)
- Research projects (FINAL_PROJECT)
- Community engagement (EXTENSION_ACTIVITIES)
- Sports & arts (PHYSICAL_EDUCATION)

Step 2: Filtering Students with No Failures or Dropouts

We will retain only students who:

- Have graduated successfully (exit_form == "Graduate").
- Have no recorded failures (situation does not include "Failed" or "Dropout").

Step 3: Defining Academic Success

We will use final_average as the measure of academic success.

Planned Analyses

1. Correlation Tests (Pearson & Spearman):
 - Pearson measures linear relationships.
 - Spearman captures rank-based (monotonic) relationships.
2. Regression Model:
 - Predicts academic success using extra-curricular participation, controlling for factors like gender

1. **Control Variables:** Extract using extra-curricular participation, controlling for admission age, gender, admission age, and quota type.
3. **Propensity Score Matching (PSM):**
- Matches students with and without extra-curricular participation based on similar characteristics to estimate causal impact.

Now, I'll extract students who participated in extra-curricular activities and those who did not, ensuring they meet the no failure/dropout criteria.

Dataset Summary

- 578 students who successfully graduated (no failures or dropouts).
- Extra-curricular participation** (`extra_curricular`):
 - `1` = Participated in at least one extra-curricular activity.
 - `0` = Did not participate.
- Final Average** (`final_average`): Measure of academic success.
- Control Variables:**
 - `gender`
 - `admission_age`
 - `quota_type` (type of admission quota used)

Now, I'll proceed with correlation analysis (Pearson & Spearman). [↩]

Correlation Analysis Results

- Pearson correlation: -0.003 ($p = 0.942$) → No significant linear relationship.
- Spearman correlation: -0.003 ($p = 0.952$) → No significant monotonic relationship.

This suggests that extra-curricular participation has no clear correlation with final academic performance. However, correlation does not imply causation, so we will proceed with regression analysis to control for other factors. [↩]

Envie uma mensagem para o ChatGPT

 Anexar

 Voz

O ChatGPT pode cometer erros. Considere verificar informações importantes.

