



Entrar

Cadastrar

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 will first inspect the dataset to understand its structure and then proceed with cleaning it according to your specifications. Let me load and preview the data.

The dataset contains 30,130 entries with 34 columns, including student information, admission details, dropout data, academic performance, and subject-related information.

Now, I will clean the dataset based on your criteria:

1. Keep only records where admission\_period is between 2006.1 and 2019.2.
2. Remove rows where enrollment\_type is "Exempted".
3. Exclude entries where situation is "In Progress" ↓

Let's see the result:

The cleaned dataset now contains 16,563 records. It is stored in df\_filtered and is ready for analysis.

Please provide a specific analytical question you'd like me to answer using this dataset. [?]

Evaluate the academic performance (e.g., average grade, improvement rate) of the bottom 10% of students in mathematics sector courses. Propose three analytical methods and select the best option.

To evaluate the academic performance of the bottom 10% of students in Mathematics sector courses, I will follow these steps:

#### Step 1: Define Performance Metrics

We will assess students' academic performance using:

1. **Final Average Grade** – A direct measure of performance in courses.
2. **Improvement Rate** – Change in performance over time (if multiple records exist per student).

#### Step 2: Filter Mathematics Courses & Identify Bottom 10%

We will:

- Filter df\_filtered to include only courses from the Mathematics sector.
- Identify the bottom 10% of students based on the final average grade. ↓

#### Step 3: Analytical Methods

We will compare three different methods for evaluation:

1. **Descriptive Statistics Approach**
  - Compute mean, median, and distribution of grades for the bottom 10%.
  - Compare with overall mathematics students.
  - Pros: Simple and interpretable.
  - Cons: Doesn't account for performance changes over time.
2. **Time-Series Analysis of Improvement**
  - Track students' grade trends over multiple semesters.
  - Calculate the **grade change rate: Improvement Rate** =  $\frac{\text{Final Grade} - \text{Initial Grade}}{\text{Initial Grade}}$ .
  - Pros: Measures progress over time.
  - Cons: Requires multiple grade records per student.
3. **Clustering-Based Performance Segmentation**
  - Apply **K-Means Clustering** to group students based on grades and trends.
  - Helps differentiate **persistent low performers** from **improving students**.
  - Pros: Identifies patterns in struggling students.
  - Cons: Computationally intensive, requires careful cluster tuning. ↓

#### Step 4: Best Method Selection

Since the goal is to evaluate both overall performance and improvement trends, I will use **Method 2 (Time-Series Analysis of Improvement)** as it best captures whether struggling students improve.

Now, I will execute the analysis.

#### Analysis Results

- **Average Final Grade (Bottom 10%):** 0.41 (on a 0-10 scale, indicating very low performance).
- **Average Improvement Rate:** ∞ (infinity) due to some students having an initial grade of 0, causing division errors.
- **Total Students Analyzed:** 206 in Mathematics sector courses.

#### Next Steps to Refine Analysis

To handle the infinite improvement rate, I will:

1. Exclude cases where the previous grade is 0 (to avoid division by zero).
2. Recalculate the improvement rate.

Let me refine the analysis. [?]

#### Refined Analysis Results

- **Average Improvement Rate (excluding division errors):** 42.77%

- This suggests that, on average, struggling students improved their grades by **42.77%** over time.
- **Average Final Grade: 0.41** (very low, confirming poor performance).

### Conclusion

- While these students start with **extremely low grades**, those who continue tend to **show significant improvement** over time.
- However, the final performance remains low, suggesting **systemic learning challenges** in mathematics courses.

Would you like a deeper breakdown, such as improvement trends over semesters or comparisons with top-performing students? [\[+\]](#)

Envie uma mensagem para o ChatGPT

 Anexar

 Voz

O ChatGPT pode cometer erros. Considere verificar informações importantes.