

1.3 - pred

17 de julho de 2024



historico_estratificado.csv
Planilha

Analise os dados e limpe as colunas da seguinte forma:
período_ingresso: de 2006.1 até 2019.2
tipo_matricula: remover "Dispensa"
situacao: remover "Em Curso"
o novo dataframe após a limpeza deve se chamar df_filtrado



Primeiro, vou carregar e examinar o conteúdo do arquivo CSV para entender melhor a estrutura dos dados. Vamos começar com isso.

Agora que visualizamos os dados, vamos filtrar o DataFrame `df` de acordo com os critérios que você especificou:

1. **período_ingresso**: Manter apenas os registros do período de 2006.1 até 2019.2.
2. **tipo_matricula**: Remover registros com o valor "Dispensa".
3. **situacao**: Remover registros com o valor "Em Curso".

Vou proceder com essas filtrações e criar o novo DataFrame `df_filtrado`. [↵]

O DataFrame `df_filtrado` foi criado com sucesso, contendo apenas os registros que atendem aos critérios especificados:

- Período de ingresso entre 2006.1 e 2019.2.
- Tipo de matrícula diferente de "Dispensa".
- Situação diferente de "Em Curso".

Atualmente, `df_filtrado` possui 13,831 registros. Se precisar de mais alguma manipulação de dados ou análise, estou à disposição para ajudar! [↵]

É possível classificar alunos em categorias de desempenho acadêmico (por exemplo, alto, médio, baixo) com base em suas médias finais e carga horária de disciplinas? Defina 3 opções de como solucionar essa questão e siga a melhor



Para classificar alunos em categorias de desempenho acadêmico com base nas médias finais e carga horária das disciplinas, podemos considerar algumas abordagens:

Opção 1: Cálculo de Pontuação Ponderada

- **Descrição**: Calcular uma pontuação ponderada que considera tanto a média final quanto a carga horária das disciplinas. A fórmula poderia ser algo como $Pontuação = Média\ Final \times Carga\ Horária$. Em seguida, definiríamos limites para categorizar as pontuações em alto, médio e baixo.
- **Vantagens**: Utiliza tanto a média quanto a carga horária, refletindo o esforço em disciplinas com maior carga horária.
- **Desvantagens**: A ponderação pode precisar de ajustes para balancear a influência de cada componente.

Opção 2: Classificação Baseada na Média Final

- **Descrição**: Simplesmente usar a média final para classificar o desempenho, ignorando a carga horária. Definiríamos limites específicos para as médias (por exemplo, acima de 7 é alto, entre 5 e 7 é médio, abaixo de 5 é baixo).
- **Vantagens**: Simplicidade e fácil interpretação.
- **Desvantagens**: Ignora a carga horária, que pode ser um fator importante no desempenho dos alunos em disciplinas mais exigentes.

Opção 3: Análise Multivariada

- **Descrição:** Utilizar métodos de análise multivariada, como análise de cluster, para classificar os alunos com base em múltiplas variáveis (média final e carga horária).
- **Vantagens:** Abordagem baseada em dados que pode revelar padrões não evidentes em métodos mais simples.
- **Desvantagens:** Requer conhecimento mais avançado em estatística e pode ser mais difícil de explicar os resultados aos interessados.

Decisão e Implementação

Dentre as opções, a **Opção 1** parece ser a mais equilibrada, pois leva em conta tanto o desempenho quanto a carga de trabalho do aluno, fornecendo uma visão mais holística. Vamos seguir com essa abordagem para classificar os alunos em categorias de desempenho acadêmico. Primeiro, precisamos calcular a pontuação ponderada para cada aluno e em seguida categorizar essas pontuações.

Os alunos foram classificados em categorias de desempenho acadêmico com base na pontuação ponderada, que é o produto da média final pela carga horária. As categorias foram definidas usando os percentis 33 e 66 da distribuição de pontuações ponderadas:

- **Baixo:** Pontuação até 306.0
- **Médio:** Pontuação entre 306.0 e 462.0
- **Alto:** Pontuação acima de 462.0

Aqui estão alguns exemplos do resultado:

Média Final	Carga Horária	Pontuação Ponderada	Categoria Desempenho
7.5	300	2250.0	Alto
5.0	60	300.0	Baixo
8.3	60	498.0	Alto
1.7	60	102.0	Baixo
7.9	60	474.0	Alto

Se precisar de mais detalhes, ajustes ou outras análises, fique à vontade para solicitar! [↗](#)

Comece com Data Analyst

[Denunciar conteúdo](#) | [Termos de uso](#) | [Política de privacidade](#) |

?