

1.1 - 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 os dados do arquivo CSV para entender sua estrutura. Depois disso, posso prosseguir com as etapas de limpeza conforme as instruções fornecidas. Vamos começar lendo o arquivo.

O dataframe contém várias colunas, mas vamos focar nas especificadas para a limpeza:

1. **período_ingresso:** Filtraremos para manter apenas os valores de 2006.1 a 2019.2.
2. **tipo_matricula:** Removeremos as linhas onde o valor é "Dispensa".
3. **situacao:** Removeremos as linhas onde o valor é "Em Curso".

Agora, vou realizar essas operações de limpeza para criar o `df_filtrado`. [↗]

A limpeza dos dados foi realizada com sucesso e resultou em um dataframe chamado `df_filtrado`, contendo 13.831 entradas. As etapas executadas foram:

1. Filtragem por período de ingresso, mantendo apenas de 2006.1 a 2019.2.
2. Remoção de linhas com o tipo de matrícula "Dispensa".
3. Remoção de linhas com a situação "Em Curso".

Se você precisar de mais análises ou ajustes nesses dados, estou à disposição para ajudar! [↗]

Considere que a média geral do aluno é definida como a média de todas as notas não nulas nas disciplinas que o aluno cursou. A idade de ingresso e idade de evasão do aluno tem alguma influência significativa na sua média geral? Defina 3 opções de como solucionar essa questão e siga a melhor



Para avaliar se a idade de ingresso e a idade de evasão influenciam significativamente a média geral do aluno, podemos considerar as seguintes abordagens:

1. **Correlação Estatística:** Avaliar a correlação entre a média geral do aluno e suas idades de ingresso e de evasão. Isso pode ser feito usando coeficientes de correlação, como o coeficiente de Pearson, que mede a força e a direção da relação linear entre duas variáveis. Valores próximos de 1 ou -1 indicam uma forte correlação positiva ou negativa, respectivamente.
2. **Modelagem Preditiva:** Construir um modelo de regressão linear onde a média geral é a variável dependente, e as idades de ingresso e evasão são as variáveis independentes. Isso permitirá não apenas verificar a influência (através dos coeficientes do modelo) de cada idade na média geral, mas também avaliar o quão bem essas variáveis explicam a variância na média geral.
3. **Análise Visual:** Criar gráficos de dispersão (scatter plots) entre a média geral e cada uma das idades (ingresso e evasão) e incluir uma linha de tendência. Isso ajudaria a visualizar se há alguma tendência ou padrão claro que sugira uma relação entre essas variáveis.

Melhor Abordagem: Modelo de Regressão Linear

Optar por construir um modelo de regressão linear pode ser a abordagem mais informativa porque:

- **Quantificação de Influência:** Fornece uma quantificação exata de quanto cada idade impacta na média geral, através dos coeficientes do modelo.
- **Avaliação de Significância:** Permite testar a significância estatística de cada variável no modelo, dando uma noção de quão confiáveis são os resultados.

- **Explicação de Variância:** Fornece uma medida (R^2) de quanto da variabilidade na média geral é explicada pelas idades de ingresso e evasão.

Vamos implementar esta abordagem e analisar os resultados. Primeiro, precisamos preparar os dados, garantindo que não há valores nulos nos campos relevantes e então ajustar o modelo de regressão linear.

Os resultados da análise com o modelo de regressão linear são os seguintes:

- O **coeficiente de determinação (R^2)** é aproximadamente 0.00018. Este valor indica que as idades de ingresso e evasão explicam apenas uma fração muito pequena da variabilidade nas médias gerais dos alunos. Em termos práticos, essas variáveis têm um impacto quase nulo na previsão da média geral dos alunos.
- Os **coeficientes** para idade de ingresso e idade de evasão são, respectivamente, aproximadamente 3.13 e -3.57. Isso sugere que um aumento na idade de ingresso tende a ter um pequeno efeito positivo na média geral, enquanto um aumento na idade de evasão tende a ter um pequeno efeito negativo. No entanto, a magnitude desses coeficientes é muito baixa, o que implica que as mudanças na média geral associadas a mudanças nessas idades são mínimas.

Conclusão

Baseado na análise, podemos concluir que as idades de ingresso e evasão do aluno não têm influência significativa na média geral dos alunos, tanto em termos de explicação da variabilidade quanto em termos de magnitude de impacto.

Se houver outras questões ou uma direção diferente que você gostaria de explorar, sinta-se à vontade para informar! [\[>-\]](#)

Comece com Data Analyst

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

?