

2.3 - diag

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 analisar os dados do arquivo que você forneceu. Isso me ajudará a entender a estrutura dos dados e a realizar as limpezas necessárias. Vamos começar por verificar as primeiras linhas do arquivo para identificar as colunas e os tipos de dados.

O arquivo contém diversas colunas. Vou agora filtrar e limpar os dados conforme suas instruções:

1. Manter registros no `período_ingresso` entre 2006.1 e 2019.2.
2. Remover linhas onde `tipo_matricula` é igual a "Dispensa".
3. Excluir linhas onde `situacao` é "Em Curso".

Vou prosseguir com esses passos. [↵]

A limpeza dos dados foi concluída com sucesso. O dataframe `df_filtrado` agora contém 13,831 registros que atendem aos critérios especificados:

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

Se precisar de mais alguma operação com esses dados ou de outra análise, é só avisar! [↵]

Quão consistentes são as notas dos alunos ao longo do tempo?



Para analisar a consistência das notas dos alunos ao longo do tempo, podemos utilizar várias abordagens estatísticas. Uma opção é calcular a variabilidade das médias finais de cada aluno ao longo de diferentes períodos, observando indicadores como a média, desvio padrão e possivelmente a variação do coeficiente. Também podemos examinar graficamente as tendências ou padrões das notas ao longo do tempo.

Vou começar calculando esses indicadores estatísticos para ver a consistência das notas. Vamos lá?

Aqui estão algumas estatísticas descritivas sobre a consistência das notas dos alunos ao longo do tempo:

- **Média das notas:** 6.34 (em uma escala provável de 0 a 10)
- **Desvio padrão médio das notas por aluno:** 1.83, o que indica uma variabilidade moderada.
- **Coeficiente de variação (CV):** Em média, é 0.36. O coeficiente de variação é uma medida relativa de dispersão e sugere que, em média, há uma variabilidade de 36% em torno da média de notas de cada aluno.
- **Distribuição de CV:** O CV varia de 0 (nenhuma variação nas notas) a aproximadamente 2.24, com 25% dos alunos tendo um CV até 0.137 e 75% dos alunos até 0.493. Isso mostra que a maioria dos alunos tem uma variação relativamente baixa a moderada nas notas ao longo do tempo, mas também há casos de variação alta.

Essas estatísticas indicam que há uma certa inconsistência nas notas dos alunos ao longo do tempo, mas para a maioria, essa variação não é extremamente alta. Se desejar, podemos explorar visualmente esses dados ou fazer uma análise mais específica. [↵]

