

2.2 - desc

17 de julho de 2024

ⓘ Este chat contém arquivos ou imagens produzidos por Análise de Dados Avançada que ainda não estão visíveis nos Chats Compartilhados.



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



Vou começar lendo o arquivo que você enviou para entender como os dados estão estruturados e então aplicar os filtros conforme solicitado.

O dataset contém 34 colunas. Vamos aplicar os filtros solicitados:

1. **período_ingresso:** Limitar os dados de 2006.1 até 2019.2.
2. **tipo_matricula:** Remover entradas com o valor "Dispensa".
3. **situacao:** Remover entradas com o valor "Em Curso".

Vou começar pela filtragem dos períodos de ingresso, depois remover as entradas específicas nas colunas de tipo de matrícula e situação. [-]

O novo dataframe, `df_filtrado`, foi criado aplicando os filtros solicitados:

- Período de ingresso limitado entre 2006.1 e 2019.2.
- Entradas com o tipo de matrícula "Dispensa" foram removidas.
- Entradas com a situação "Em Curso" também foram removidas.

Se precisar de mais alguma análise ou manipulação desses dados, é só me avisar! [-]

Qual é a distribuição de notas pelo período da matrícula?



Vamos começar analisando a distribuição das notas (média final) pelo período de matrícula. Vou criar um gráfico que mostra a média das notas por cada período de matrícula, bem como a dispersão das notas para cada período.



O resultado de imagem não é suportado no chat compartilhado

O gráfico acima mostra a distribuição das médias finais dos alunos por período de matrícula. Cada caixa no gráfico de caixas (boxplot) representa a distribuição das notas para um determinado período, onde:

- A linha central da caixa indica a mediana das notas.
- As bordas da caixa mostram o primeiro e terceiro quartil.
- Os pontos fora das "linhas de bigode" são considerados valores atípicos ou outliers.

Observa-se que a distribuição das notas varia um pouco de um período para outro, mas em geral, as medianas estão em torno de 6 a 8, indicando uma tendência de notas médias nessa faixa.

Se precisar de mais alguma análise específica ou um detalhamento maior de um período em particular, me avise! [-]

Comece com Data Analyst

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

