

3.2 - desc

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



Vamos começar analisando o conteúdo do arquivo CSV que você enviou. Primeiro, vou carregar os dados e dar uma olhada nas primeiras linhas para entender a estrutura do arquivo. Vamos lá!

Agora vamos filtrar os dados com base nos critérios que você especificou:

1. **Período de Ingresso:** De 2006.1 até 2019.2
2. **Tipo de Matrícula:** Remover registros com "Dispensa"
3. **Situação:** Remover registros com "Em Curso"

Primeiro, verificarei os valores únicos dessas colunas para garantir que o filtro possa ser aplicado corretamente. [↗]

Os valores únicos confirmam a presença dos critérios que você deseja filtrar. Agora, vamos proceder com a filtragem dos dados:

1. **Período de Ingresso:** Selecionaremos dados de 2006.1 até 2019.2.
2. **Tipo de Matrícula:** Excluiremos registros com "Dispensa".
3. **Situação:** Excluiremos registros com "Em Curso".

Vou criar o novo DataFrame conforme as especificações e chamá-lo de `df_filtrado`. [↗]

A filtragem foi realizada com sucesso, e o novo DataFrame `df_filtrado` contém 13.831 registros, cada um atendendo 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 análises ou ajustes nesses dados, é só me dizer! [↗]

Como a entropia da distribuição de alunos por setor acadêmico mudou ao longo dos últimos 5 períodos?



Para calcular a entropia da distribuição de alunos por setor acadêmico ao longo dos últimos 5 períodos, primeiro precisamos entender a distribuição de alunos por setor em cada um desses períodos. Então, precisamos calcular a entropia para cada distribuição.

Vou começar identificando os últimos 5 períodos únicos no DataFrame `df_filtrado` e, em seguida, calcular a distribuição e a entropia para cada um desses períodos.

As entropias da distribuição de alunos por setor acadêmico nos últimos 5 períodos são as seguintes:

- **2018.2:** Entropia de aproximadamente 0.748
- **2019.0:** Entropia de 0.0 (indicando uma distribuição provavelmente com todos os alunos em um único setor)
- **2019.1:** Entropia de aproximadamente 0.745
- **2019.2:** Entropia de aproximadamente 0.683
- **2020.1:** Entropia de 0.0 (indicando, assim como em 2019.0, uma distribuição concentrada)

Uma entropia de 0 indica que não há diversidade na amostra, ou seja, todos os alunos pertencem a

um único setor. As outras entropias, variando entre 0.683 e 0.748, sugerem uma diversidade moderada na distribuição de alunos entre os setores acadêmicos.

Se precisar de mais detalhes ou de outra análise, me avise! [\[~\]](#)

Comece com Data Analyst

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

?