

2.1 - 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



Primeiro, vou carregar e examinar os dados do arquivo `historico_estratificado.csv` para entender sua estrutura. Vamos começar lendo o arquivo e visualizando as primeiras linhas do dataframe.

O dataset contém várias colunas, mas para a limpeza solicitada, vamos focar nas seguintes colunas:

- `período_ingresso`
- `tipo_matricula`
- `situacao`

Vamos agora filtrar os dados de acordo com as condições especificadas:

1. `período_ingresso` deve estar entre 2006.1 e 2019.2.
2. `tipo_matricula` não deve conter a entrada "Dispensa".
3. `situacao` não deve conter a entrada "Em Curso".

Após aplicar esses filtros, armazenaremos o resultado no novo dataframe `df_filtrado`. Vamos fazer isso agora. [↗]

A limpeza dos dados foi realizada com sucesso. O novo dataframe `df_filtrado` contém 13.831 registros e 34 colunas. Foram aplicados os seguintes filtros:

- O período de ingresso está entre 2006.1 e 2019.2.
- Foram removidas as entradas com o tipo de matrícula "Dispensa".
- Foram removidas as entradas com a situação "Em Curso".

Se precisar de mais alguma análise ou modificação nesses dados, fique à vontade para me pedir! [↗]

Qual é a proporção de alunos evadidos por ano de ingresso?



Para calcular a proporção de alunos evadidos por ano de ingresso, precisamos primeiro definir o que consideramos como "evadido" no contexto dos dados. Vamos examinar a coluna `forma_saida` para identificar as categorias que indicam evasão e, em seguida, calcular a proporção de alunos evadidos por ano de ingresso.

Vou começar analisando os valores únicos na coluna `forma_saida` para identificar os possíveis indicadores de evasão.

A coluna `forma_saida` inclui as seguintes categorias:

- Graduado
- Evadido
- Reingresso
- Regular
- Outros

Para este cálculo, consideraremos "Evadido" como a indicação de que o aluno evadiu. Agora, vou calcular a proporção de alunos evadidos por ano de ingresso. Vamos primeiro extrair o ano de ingresso da coluna `período_ingresso`, e depois agrupar os dados para calcular a proporção. [↗]

Aqui está a proporção de alunos evadidos por ano de ingresso:

- 2006: 9.09%
- 2007: 20.75%
- 2008: 15.23%
- 2009: 26.34%
- 2010: 25.84%
- 2011: 37.27%
- 2012: 34.70%
- 2013: 37.15%
- 2014: 26.21%
- 2015: 20.99%
- 2016: 22.72%
- 2017: 19.31%
- 2018: 11.01%
- 2019: 5.26%

Observamos que a proporção de evasão varia significativamente ao longo dos anos, com uma tendência de decréscimo a partir de 2017. Se precisar de mais detalhes ou análises adicionais, estou à disposição! [\[↗\]](#)

Comece com Data Analyst

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

?