

2.1 - 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 para entender sua estrutura. Em seguida, realizarei as limpezas especificadas nas colunas ``período_ingresso``, ``tipo_matricula`` e ``situacao``. Vamos começar!

O dataframe possui várias colunas, incluindo as que nos interessam para a limpeza:
``período_ingresso``, ``tipo_matricula`` e ``situacao``.

Agora, vou proceder com os filtros conforme suas instruções:

1. Manter apenas as linhas onde ``período_ingresso`` está entre 2006.1 e 2019.2.
2. Remover linhas onde ``tipo_matricula`` é igual a "Dispensa".
3. Remover linhas onde ``situacao`` é igual a "Em Curso".

Vou aplicar esses filtros agora. [↗]

A limpeza dos dados foi concluída com sucesso. O novo dataframe ``df_filtrado`` contém agora 13.831 linhas e 34 colunas, após aplicarmos os seguintes filtros:

- Manutenção apenas dos registros de ``período_ingresso`` entre 2006.1 e 2019.2.
- Remoção de registros onde ``tipo_matricula`` é "Dispensa".
- Remoção de registros onde ``situacao`` é "Em Curso".

Se precisar de mais alguma análise ou ajuste nesses dados, estou à disposição para ajudar! [↗]

Determine se a forma de ingresso tem um impacto significativo nas taxas de graduação e evasão



Para investigar se a forma de ingresso tem um impacto significativo nas taxas de graduação e evasão, podemos começar analisando a distribuição dos alunos que se graduaram ou evadiram em relação às diferentes formas de ingresso. Vamos primeiro criar uma tabela cruzada que mostra a quantidade de alunos graduados e evadidos por cada forma de ingresso. Depois, podemos aplicar um teste estatístico para ver se as diferenças observadas são estatisticamente significativas.

Primeiramente, vou criar a tabela cruzada usando os dados. Considerarei como "graduados" aqueles cuja ``forma_saida`` é "Graduado" e como "evadidos" aqueles cuja ``forma_saida`` contém palavras chave indicativas de evasão (como "Evadido", "Cancelamento", "Reprovou", entre outros).

Aqui está a tabela cruzada que mostra o número de graduados e evadidos para cada forma de ingresso. Podemos ver que algumas formas de ingresso como "VESTIBULAR" e "SISU" têm números altos tanto para graduados quanto para evadidos, enquanto outras formas têm poucas observações, o que pode dificultar a análise estatística devido à baixa frequência.

Para determinar se as diferenças entre as formas de ingresso são estatisticamente significativas em termos de taxas de graduação e evasão, podemos aplicar o teste Qui-quadrado de independência. Este teste avalia se as frequências observadas em categorias específicas são diferentes das frequências esperadas, indicando uma possível associação entre as variáveis. Vou aplicar este teste agora. [↗]

O teste Qui-quadrado de independência resultou em um valor p de aproximadamente 0.0, o que indica que há uma diferença estatisticamente significativa entre as formas de ingresso no que diz respeito às taxas de graduação e evasão. Com um valor p muito pequeno, podemos rejeitar a

hipótese nula de que as formas de ingresso e os status de graduação/evasão são independentes.

Isso sugere que a forma de ingresso pode de fato ter um impacto significativo na probabilidade de um estudante se graduar ou evadir. Para políticas educacionais e de planejamento, esses insights podem ser bastante úteis na adaptação ou revisão das estratégias de admissão.

Se precisar de mais análises ou detalhes específicos sobre estes resultados, fique à vontade para pedir! [\[>-\]](#)

Comece com Data Analyst

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

?