

3.1 - 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



Para prosseguir com a análise e limpeza dos dados do arquivo "historico_estratificado.csv", primeiramente vou carregar e examinar os dados para entender melhor sua estrutura e conteúdo. Vou começar lendo o arquivo e mostrando as primeiras linhas, além de verificar os tipos de dados das colunas.

Agora que temos uma visão geral dos dados, vamos proceder com as etapas de limpeza solicitadas:

1. **Filtrar `período_ingresso`:** Manteremos apenas os registros com valores de 2006.1 até 2019.2.
2. **Remover registros com `tipo_matricula` igual a "Dispensa".**
3. **Excluir registros onde a `situacao` seja "Em Curso".**

Vou realizar essas etapas e criar o novo DataFrame chamado `df_filtrado`. [↗]

A limpeza foi concluída com sucesso. O novo DataFrame `df_filtrado` agora contém 13.831 registros, após a remoção de registros fora do período de ingresso especificado, registros com o tipo de matrícula "Dispensa" e situações "Em Curso".

Caso precise realizar mais alguma análise ou ajuste nos dados, estou à disposição para ajudar! [↗]

Considere que a média geral do aluno é definida como a média de todas as notas não nulas nas disciplinas que o aluno cursou. Além disso, considere que a cada ano existem dois períodos, o primeiro período do ano é caracterizado por terminar com ".1" e o segundo período do ano termina com ".2", por exemplo, no ano de 2015 existem os períodos 2015.1 e 2015.2. A partir destas definições, qual é a correlação entre a média geral dos alunos e ingressar em cada período do ano (primeiro e segundo)?



Para calcular a correlação entre a média geral dos alunos e o período de ingresso no ano (se é o primeiro ou o segundo semestre), vamos seguir os seguintes passos:

1. **Calcular a média geral de cada aluno:** Utilizaremos a coluna `media_final` para calcular a média de todas as disciplinas cursadas por cada aluno, considerando apenas notas não nulas.
2. **Determinar o período de ingresso:** Criaremos uma nova coluna para indicar se o ingresso ocorreu no primeiro ou no segundo período do ano.
3. **Calcular a correlação:** Verificar a relação entre essas duas variáveis.

Primeiramente, vamos calcular a média geral de cada aluno. Em seguida, criaremos a coluna para indicar o período de ingresso e, por fim, calcularemos a correlação. Vamos começar com o cálculo da média geral para cada aluno.

A média geral foi calculada para cada aluno e adicionada ao DataFrame. Agora, vamos criar uma nova coluna para indicar se o ingresso foi no primeiro período (".1") ou no segundo período (".2") do ano. Depois disso, calcularemos a correlação entre a média geral e esse indicador de período de ingresso. [↗]

A correlação entre a média geral dos alunos e o período de ingresso no ano (primeiro ou segundo semestre) é de aproximadamente -0.177. Essa correlação negativa sugere uma relação fraca, indicando que alunos que ingressam no segundo semestre tendem a ter uma média geral ligeiramente inferior em comparação com aqueles que ingressam no primeiro semestre.

Se precisar de mais análises ou tiver outras perguntas, estou aqui para ajudar! [↗]

Comece com Data Analyst

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

?