

Desafio de Dados

Aspirante: Ádson Arantes Viana

Escolhi utilizar a linguagem Python e a biblioteca Pandas para manipular os dados fornecidos. Então, fiz a instalação da biblioteca e a importação no arquivo de python. Importei também o arquivo da base despadronizada e a armazenei em `alunos_df`.

Para o objetivo de padronizar a coluna “sexo”, a primeira coisa feita foi verificar quais eram todas as variações de respostas para a coluna “sexo” e para isso utilizei o `.value_counts()`. Dessa forma, encontrei que todas as variações existentes são: M, fem, F, Feminino, masc e Masculino. Para padronizar a coluna substituí todos os seus elementos com base nas variações de respostas, buscando todos os elementos das linhas da coluna “sexo” que está no dataframe chamado `alunos_df`. Se o elemento da linha da coluna “sexo” for “M” ou “masc” ele será substituído por “Masculino”, semelhantemente ocorre com “F” e “fem” que são substituídos por “Feminino”.

Para facilitar as manipulações, em seguida criei e preenchi as colunas Médias e Aprovado. Para isso, precisei padronizar os valores das colunas de nota com o uso do ponto e transformá-los em valores do tipo float. A partir disso, criei a coluna com as médias dos alunos com base na fórmula fornecida e também construí o comando para analisar se o aluno está aprovado ou não.

Por fim, realizei as formatações necessárias para padronizar as colunas `notas_matematica`, `notas_portugues`, Médias e frequência de acordo com a formatação brasileira. Dessa forma, todos os objetivos do desafio foram alcançados.