**Descrição Geral:**

Este código realiza a limpeza, transformação e análise de dados do ENEM 2020 utilizando bibliotecas como pandas, numpy, matplotlib e seaborn. O foco principal é analisar o desempenho dos alunos com base em variáveis como sexo, faixa etária, escolaridade, presença nas provas e notas. A análise busca extrair insights sobre as relações entre essas variáveis e o desempenho acadêmico dos participantes.

**Passos do Código:**

1. **Imports e Configurações Iniciais:**
    - **Bibliotecas:**
        - pandas, numpy, matplotlib.pyplot, seaborn: Usadas para manipulação de dados, análise estatística e visualização.
        - warnings: Para suprimir mensagens de aviso durante a execução.
        - os, sys: Utilizadas para manipulação de arquivos e caminhos no sistema operacional.
    - **Configuração de exibição do pandas:** O código é configurado para exibir todas as colunas e linhas do dataframe sem truncamentos.
2. **Leitura dos Dados:**
    - O dataset é carregado de um arquivo CSV com a função pd.read_csv().
    - Valores ausentes são identificados e tratados, substituindo palavras-chave como "n/a", "na" e "undefined" por NaN.
3. **Identificação e Cálculo de Valores Ausentes:**
    - A função func_calc_percentual_valores_ausentes() é usada para calcular a porcentagem de valores ausentes no dataset.
    - Uma tabela é gerada com a porcentagem de valores ausentes por coluna. Colunas com mais de 60% de dados ausentes são removidas, exceto aquelas com informações críticas como TP_ENSINO e TP_DEPENDENCIA_ADM_ESC.
4. **Tratamento de Valores Ausentes e Esquema de Dados:**
    - Após a remoção de colunas com muitos valores ausentes, a função func_calc_percentual_valores_ausentes_coluna() é executada novamente para as colunas restantes.
    - Algumas variáveis, como TP_ENSINO e TP_DEPENDENCIA_ADM_ESC, são analisadas quanto à distribuição para identificar possíveis desvios.

5. **Seleção e Limpeza de Colunas Importantes:**
    - São definidas as colunas de interesse e um novo dataframe é criado, contendo apenas essas variáveis.
    - Algumas variáveis categóricas são convertidas para o tipo object para facilitar a manipulação de dados textuais.
6. **Mapeamento e Substituição de Códigos Categóricos:**
    - Dicionários de mapeamento são definidos para transformar códigos numéricos em descrições mais legíveis (ex.: "Solteiro(a)", "Pública").
    - A transformação é realizada utilizando a função .map().
7. **Renomeação de Colunas:**
    - As colunas do dataframe são renomeadas para nomes mais claros e padronizados, como por exemplo, NU_INSCRICAO para COD_INSCRICAO.
8. **Cálculo de Soma e Média das Notas:**
    - Uma nova coluna SOMA_NOTAS é criada para armazenar a soma das notas de quatro disciplinas (Ciências da Natureza, Ciências Humanas, Linguagem e Código, Matemática).
    - A coluna MEDIA_NOTAS é calculada como a média das notas dessas disciplinas.
    - A média das notas por escola é calculada e a escola com a maior média é identificada.
9. **Visualização de Dados:**
    - Um gráfico de barras é gerado para mostrar a média das notas por tipo de escola (Pública, Privada, Exterior, etc.).
    - O gráfico é ajustado para mostrar etiquetas com a média das notas e as legendas são rotacionadas para melhorar a visualização.
10. **Análise de Aluno com Melhor Desempenho:**
    - O aluno com a maior média de notas é identificado, com seu número de inscrição (ou código) e a média de notas sendo exibidos.
11. **Cálculo da Média Geral:**
    - A média geral das notas de todos os alunos é calculada e impressa para proporcionar uma visão geral do desempenho.

**Funções Personalizadas:**

- **func_calc_percentual_valores_ausentes(df)**: Calcula a porcentagem de valores ausentes por coluna no dataframe.
- **func_calc_percentual_valores_ausentes_coluna(df)**: Retorna uma tabela com a porcentagem de valores ausentes por coluna.
- **convert_to_string(df, columns)**: Converte as colunas especificadas de um dataframe para o tipo string.
- **Trataoutlier**: é uma classe onde trata dos outlier das variáveis numéricas.

**Principais Variáveis e Dicionários:**

- **Dicionários de Mapeamento:**
    - faixa_etaria_dicionario: Mapeia códigos de faixa etária para descrições como "18-24 anos", "25-34 anos", etc.
    - cor_raca_dicionario: Mapeia códigos de cor/raça para descrições como "Branca", "Negra", etc.
    - escola_dicionario: Mapeia os códigos das escolas para categorias como "Pública" ou "Privada".
    - Outros dicionários mapeiam variáveis como estado civil, presença nas provas, nacionalidade, e status da redação.
- **Principais Variáveis:**
    - **dataset**: Contém o conjunto de dados limpo e transformado.
    - **media_por_escola**: Calcula a média das notas por tipo de escola.

**Insights:**

1. **Análise de Notas**: A relação entre o tipo de escola e a média das notas pode oferecer insights sobre o impacto do contexto educacional no desempenho.
2. **Distribuição Etária**: A faixa etária dos alunos pode ser analisada para verificar se existe uma correlação entre idade e desempenho nas provas.
3. **Desempenho por Sexo e Cor/Raça**: Análises comparativas podem revelar desigualdades de desempenho baseadas em sexo e cor/raça.
4. **Impacto da Escolaridade**: A variável de escolaridade pode ajudar a entender como o nível educacional influencia o desempenho no ENEM.