**Missão Prática – Mundo 05 – Nível 03**

**Gilvan Pereira de Oliveira – 2023.01.53256-61197**

**Polo Centro – São Lourenço Da Mata – PE**

**RPG0033 – TRATANDO A IMENSIDÃO DOS DADOS – 9001 – 2025.1**

<https://github.com/GilvanPOliveira/FullStack/tree/main/Mundo05/tratandoDados>

**Missão Prática | Tratando a imensidão dos dados**

**Contextualização**

Como Analista de Dados, você recebeu, em um novo projeto, um conjunto de dados. Sua principal tarefa é tratar os dados desse conjunto a fim de que possam ser utilizados para a descoberta de conhecimento através de sua posterior análise e interpretação.

**Texto

O conteúdo gerado por IA pode estar incorreto.**

**Texto

O conteúdo gerado por IA pode estar incorreto.**

1. Atribua os dados lidos a uma variável;
2. Verifique se os dados foram importados adequadamente:
   1. Imprima as informações gerais sobre o conjunto de dados;
   2. Imprima as primeiras e últimas N linhas do arquivo.
3. Crie uma nova variável e atribua a ela uma cópia do conjunto de dados original (variável criada no passo 4);

**Texto

O conteúdo gerado por IA pode estar incorreto.**

**Tela preta com letras brancas

O conteúdo gerado por IA pode estar incorreto. Tela preta com letras brancas

O conteúdo gerado por IA pode estar incorreto.**

1. Nessa nova variável, contendo uma cópia dos dados:
   1. Substitua todos os valores nulos da coluna ‘Calories’ por 0;
   2. Imprima o conjunto de dados para verificar se a mudança acima foi aplicada com sucesso;

**Calendário

O conteúdo gerado por IA pode estar incorreto.**

1. Ainda na nova variável:
   1. Substitua os valores nulos da coluna ‘Date’ por ‘1900/01/01’;
   2. Imprima o conjunto de dados e confira se a mudança foi aplicada com sucesso;
   3. Transforme os dados da coluna ‘Date’ em datetime usando o método ‘to\_datetime’;

**Calendário

O conteúdo gerado por IA pode estar incorreto.**

1. Tendo seguido todas as instruções anteriores, ao executar o passo anterior você deverá ter encontrado um erro informando que o valor ‘1900/01/01’ não corresponde ao formato ‘%Y/%m/%d’. Para resolver esse problema:
   1. Substitua, na coluna ‘Date’, o valor ‘1900/01/01’ por ‘NaN’;
   2. Utilizando o método ‘to\_datetime’, repita o passo de transformação dos dados da coluna ‘Date’ para datetime;
   3. Imprima o conjunto de dados para verificar se as mudanças acima foram aplicadas com sucesso;

Uma imagem contendo Calendário

O conteúdo gerado por IA pode estar incorreto.

**OBS: o valor está como NaT, pois no Pandas, é atribuído para date: NaT (not a time), e não, NaN (not a number).**

1. Nesse ponto, você deverá ter esbarrado em outro erro, informando agora que o valor "20201226" não corresponde ao formato "'%Y/%m/%d'" . Você precisará, agora, na coluna ‘Date”, transformar especificamente esse valor, atualmente uma string, para o formato datetime. Para isso você deverá combinar os métodos ‘replace’ e ‘to\_datetime’;
2. Após o passo anterior, execute novamente a transformação de todos os dados da coluna ‘Date’ para o formato datetime (usando o to\_datetime). Imprima o conjunto de dados atual para verificar se todas as transformações foram executadas com sucesso;

**Uma imagem contendo Tabela

O conteúdo gerado por IA pode estar incorreto.**

1. Por fim, remova os registros contendo valores nulos. Nesse ponto, apenas a coluna ‘Date’ possui um registro que atende a essa premissa (linha 22). Logo, utilize-a como base para realizar a transformação solicitada;
2. Imprima o dataframe e verifique se todas as transformações foram executadas conforme solicitado nos passos anteriores.

**Uma imagem contendo Calendário

O conteúdo gerado por IA pode estar incorreto.**