



**UNIVERSIDADE ESTÁCIO DE SÁ
DESENVOLVIMENTO FULLSTACK**

**Mundo 05 - Nível 03
RPG0033 - Tratando a Imensidão Dos Dados**

Thiago dos Santos Moreira - Matrícula 2022080479

Belo Horizonte – MG 2024

Objetivo da Prática

1. Descrever como ler um arquivo CSV usando a biblioteca Pandas (Python);
2. Descrever como criar um subconjunto de dados a partir de um conjunto existente usando a biblioteca Pandas (Python);
3. Descrever como configurar o número máximo de linhas a serem exibidas na visualização de um conjunto de dados usando a biblioteca Pandas (Python);
4. Descrever como exibir as primeiras e últimas “N” linhas de um conjunto de dados usando a biblioteca Pandas (Python);
5. Descrever como exibir informações sobre as colunas, linhas e dados de um conjunto de dados usando a biblioteca Pandas (Python);

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. Para tal tarefa, você deverá utilizar a linguagem Python e a biblioteca Pandas. O passo-a-passo de todo o processo de tratamento dos dados é apresentado a seguir, no roteiro de prática.

Procedimentos

1. Para essa atividade você deverá, obrigatoriamente, utilizar o conjunto de dados
2. (fornecido anteriormente, na seção “Contextualização”) composto pelas colunas
3. ID;Duration;Date;Pulse;Maxpulse;Calories
4. Crie um novo arquivo/script;
5. Leia o conteúdo do CSV fornecido, atentando-se para a necessidade ou não de
6. incluir parâmetros adicionais como os relativos ao separador dos dados, a engine e
7. o encoding;

8. Atribua os dados lidos a uma variável;
9. 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.
10. Crie uma nova variável e atribua a ela uma cópia do conjunto de dados original
11. (variável criada no passo 4);
12. 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
 3. com sucesso;
13. 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
 4. 'to_datetime';
14. Tendo seguido todas as instruções anteriores, ao executar o passo anterior você
15. deverá ter encontrado um erro informando que o valor '1900/01/01' não
16. 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
 3. da coluna 'Date' para datetime;
 4. Imprima o conjunto de dados para verificar se as mudanças acima foram
 5. aplicadas com sucesso;
17. Nesse ponto, você deverá ter esbarrado em outro erro, informando agora que o
18. valor "20201226" não corresponde ao formato "%Y/%m/%d". Você precisará,
19. agora, na coluna 'Date', transformar especificamente esse valor, atualmente uma
20. string, para o formato datetime. Para isso você deverá combinar os métodos
21. 'replace' e 'to_datetime';

22. Após o passo anterior, execute novamente a transformação de todos os dados da
23. coluna 'Date' para o formato datetime (usando o `to_datetime`). Imprima o conjunto
24. de dados atual para verificar se todas as transformações foram executadas com
25. sucesso;
26. Por fim, remova os registros contendo valores nulos. Nesse ponto, apenas a coluna
27. 'Date' possui um registro que atende a essa premissa (linha 22). Logo, utilize-a
28. como base para realizar a transformação solicitada;
29. Imprima o dataframe e verifique se todas as transformações foram executadas
30. conforme solicitado nos passos anteriores.

- Resultados esperados 🌟

O resultado esperado dessa microatividade é verificar se o aluno possui conhecimentos básicos sobre python – mais precisamente sobre a biblioteca Pandas, sendo capaz de utilizá-la na leitura e manipulação de dados, realizando tarefas como a leitura de arquivos externos, a utilização de dataframes em memória, a exibição de informações e dados, assim como o tratamento/transformação dos mesmos.