



Mundo 5

Missão Prática Nível 3

429 Polo Centro – Porto Alegre – RS

Desenvolvimento Full Stack – 2023.1 – 5º Semestre Letivo

<https://github.com/Jefferson-sandoval/missao-pratica-nivel-3-mundo-5>

Objetivo da Prática:

Realizar a leitura, análise e limpeza de um arquivo .csv com dados de atividades físicas usando a biblioteca Pandas, corrigindo valores ausentes e padronizando o formato das datas.



1. Para essa atividade você deverá, obrigatoriamente, utilizar o conjunto de dados (fornecido anteriormente, na seção “Contextualização”) composto pelas colunas ID;Duration;Date;Pulse;Maxpulse;Calories

```
ID;Duration;Date;Pulse;Maxpulse;Calories
```

```
0;60;'2020/12/01';110;130;4091
1;60;'2020/12/02';117;145;4790
2;60;'2020/12/03';103;135;3400
3;45;'2020/12/04';109;175;2824
4;45;'2020/12/05';117;148;4060
5;60;'2020/12/06';102;127;3000
6;60;'2020/12/07';110;136;3740
7;450;'2020/12/08';104;134;2533
8;30;'2020/12/09';109;133;1951
9;60;'2020/12/10';98;124;2690
10;60;'2020/12/11';103;147;3293
11;60;'2020/12/12';100;120;2507
12;60;'2020/12/12';100;120;2507
13;60;'2020/12/13';106;128;3453
14;60;'2020/12/14';104;132;3793
15;60;'2020/12/15';98;123;2750
16;60;'2020/12/16';98;120;2152
17;60;'2020/12/17';100;120;3000
18;45;'2020/12/18';90;112;NaN
19;60;'2020/12/19';103;123;3230
20;45;'2020/12/20';97;125;2430 2
1;60;'2020/12/21';108;131;3642
22;45;NaN;100;119;2820
23;60;'2020/12/23';130;101;3000
24;45;'2020/12/24';105;132;2460
25;60;'2020/12/25';102;126;3345
26;60;20201226;100;120;2500
27;60;'2020/12/27';92;118;2410
28;60;'2020/12/28';103;132;NaN
29;60;'2020/12/29';100;132;2800
30;60;'2020/12/30';102;129;3803
31;60;'2020/12/31';92;115;2430
```



2. Crie um novo arquivo/script;



3. Código:

```
import pandas as pd
import numpy as np

dados = pd.read_csv('dados.csv', sep=';')

print("Informações iniciais:")
print(dados.info())
print(dados.head())
print(dados.tail())

dados_limpos = dados.copy()

dados_limpos['Calories'] = dados_limpos['Calories'].fillna(0)

dados_limpos['Date'] = dados_limpos['Date'].fillna('1900/01/01')

dados_limpos['Date'] = dados_limpos['Date'].replace('1900/01/01',
np.nan)

dados_limpos['Date'] = dados_limpos['Date'].replace('20201226',
'2020/12/26')

dados_limpos['Date'] = dados_limpos['Date'].str.replace("'", "")

dados_limpos['Date'] = pd.to_datetime(dados_limpos['Date'],
format='%Y/%m/%d', errors='coerce')

dados_limpos = dados_limpos.dropna()

print("\nDados após limpeza:")
print(dados_limpos)

dados_limpos.to_csv('dados_limpos.csv', index=False)
```