# Desenvolvimento Full Stack

**Nível 3: Tratando A Imensidão Dos Dados**

**2023.1**

**Mundo 5 Período 2025.1**

**Breno Félix de Souza**

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
2. Crie um novo arquivo/script;
3. Leia o conteúdo do CSV fornecido, atentando-se para a necessidade ou não de incluir parâmetros adicionais como os relativos ao separador dos dados, a engine e o enconding;

```
df = pd.read_csv(
   'bd.csv',
   sep=';',
   engine='python',
   encoding='utf-8'
)
```

4. Atribua os dados lidos a uma variável;

```
dados = df
```

5. Verifique se os dados foram importados adequadamente:
    a) Imprima as informações gerais sobre o conjunto de dados;

```
Informações do Gerais:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32 entries, 0 to 31
Data columns (total 6 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   ID        32 non-null     int64
 1   Duration  32 non-null     int64
 2   Date      31 non-null     object
 3   Pulse     32 non-null     int64
 4   Maxpulse  32 non-null     int64
 5   Calories  30 non-null     object
dtypes: int64(4), object(2)
memory usage: 1.6+ KB
None
```

b) Imprima as primeiras e últimas N linhas do arquivo.

```
Primeiras 5 linhas:
   ID  Duration          Date  Pulse  Maxpulse Calories
0   0        60  '2020/12/01'    110       130     4091
1   1        60  '2020/12/02'    117       145     4790
2   2        60  '2020/12/03'    103       135     3400
3   3        45  '2020/12/04'    109       175     2824
4   4        45  '2020/12/05'    117       148     4060
Últimas 5 linhas:
    ID  Duration          Date  Pulse  Maxpulse Calories
27  27        60  '2020/12/27'     92       118     2410
28  28        60  '2020/12/28'    103       132      NaN
29  29        60  '2020/12/29'    100       132     2800
30  30        60  '2020/12/30'    102       129     3803
31  31        60  '2020/12/31'     92       115     2430
```

6. Crie uma nova variável e atribua a ela uma cópia do conjunto de dados original (variável criada no passo 4);

```
dados_copia = dados.copy()
```

7. Nessa nova variável, contendo uma cópia dos dados:

   a) Substitua todos os valores nulos da coluna ‘Calories’ por 0;
   b) Imprima o conjunto de dados para verificar se a mudança acima foi aplicada com sucesso;

```
Coluna 'Calories' após preenchimento dos valores nulos com 0:
    ID  Calories
0    0    4091.0
1    1    4790.0
2    2    3400.0
3    3    2824.0
4    4    4060.0
5    5    3000.0
6    6    3740.0
7    7    2533.0
8    8    1951.0
9    9    2690.0
10  10    3293.0
11  11    2507.0
12  12    2507.0
13  13    3453.0
14  14    3793.0
15  15    2750.0
16  16    2152.0
17  17    3000.0
18  18       0.0
19  19    3230.0
20  20       0.0
21   1    3642.0
22  22    2820.0
23  23    3000.0
24  24    2460.0
25  25    3345.0
26  26    2500.0
27  27    2410.0
28  28       0.0
29  29    2800.0
30  30    3803.0
31  31    2430.0
```

8. Ainda na nova variável:
    a) Substitua os valores nulos da coluna ‘Date’ por ‘1900/01/01’;
    b) Imprima o conjunto de dados e confira se a mudança foi aplicada com sucesso;
    c) Transforme os dados da coluna ‘Date’ em datetime usando o método ‘to_datetime’;

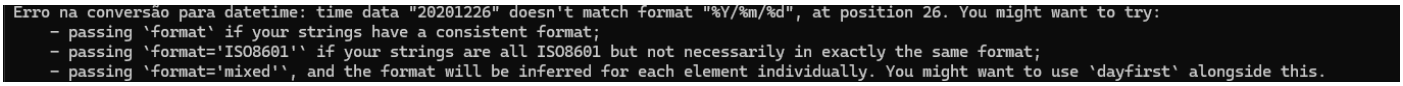
```
  dados_copia['Date'].fillna('1900/01/01', inplace=True)
Coluna 'Date' após substituição de nulos por '1900/01/01':
    ID          Date
0    0  '2020/12/01'
1    1  '2020/12/02'
2    2  '2020/12/03'
3    3  '2020/12/04'
4    4  '2020/12/05'
5    5  '2020/12/06'
6    6  '2020/12/07'
7    7  '2020/12/08'
8    8  '2020/12/09'
9    9  '2020/12/10'
10  10  '2020/12/11'
11  11  '2020/12/12'
12  12  '2020/12/12'
13  13  '2020/12/13'
14  14  '2020/12/14'
15  15  '2020/12/15'
16  16  '2020/12/16'
17  17  '2020/12/17'
18  18  '2020/12/18'
19  19  '2020/12/19'
20  20  '2020/12/20'
21   1  '2020/12/21'
22  22    1900/01/01
23  23  '2020/12/23'
24  24  '2020/12/24'
25  25  '2020/12/25'
26  26      20201226
27  27  '2020/12/27'
28  28  '2020/12/28'
29  29  '2020/12/29'
30  30  '2020/12/30'
31  31  '2020/12/31'
```

9. 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:

    a) Substitua, na coluna ‘Date’, o valor ‘1900/01/01’ por ‘NaN’;

    b) Utilizando o método ‘to_datetime’, repita o passo de transformação dos dados da coluna ‘Date’ para datetime;

    c) Imprima o conjunto de dados para verificar se as mudanças acima foram aplicadas com sucesso;




10. 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’;

```
dados_copia['Date'] =
dados_copia['Date'].replace(

  '20201226',

  pd.to_datetime('2020-12-26',
format='%Y-%m-%d')

)
```

11. 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;

```
Coluna 'Date' após todas as transformações:
    ID       Date
0    0 2020-12-01
1    1 2020-12-02
2    2 2020-12-03
3    3 2020-12-04
4    4 2020-12-05
5    5 2020-12-06
6    6 2020-12-07
7    7 2020-12-08
8    8 2020-12-09
9    9 2020-12-10
10  10 2020-12-11
11  11 2020-12-12
12  12 2020-12-12
13  13 2020-12-13
14  14 2020-12-14
15  15 2020-12-15
16  16 2020-12-16
17  17 2020-12-17
18  18 2020-12-18
19  19 2020-12-19
20  20 2020-12-20
21   1 2020-12-21
22  22        NaT
23  23 2020-12-23
24  24 2020-12-24
25  25 2020-12-25
26  26        NaT
27  27 2020-12-27
28  28 2020-12-28
29  29 2020-12-29
30  30 2020-12-30
31  31 2020-12-31
```

12. 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;

```
dados_copia.dropna(inplace=True)
```

13. Imprima o dataframe e verifique se todas as transformações foram executadas conforme solicitado nos passos anteriores.

DataFrame final após todas as transformações:

| | ID | Duration | Date | Pulse | Maxpulse | Calories |
|---|---|---|---|---|---|---|
| 0 | 0 | 60 | 2020-12-01 | 110 | 130 | 4091.0 |
| 1 | 1 | 60 | 2020-12-02 | 117 | 145 | 4790.0 |
| 2 | 2 | 60 | 2020-12-03 | 103 | 135 | 3400.0 |
| 3 | 3 | 45 | 2020-12-04 | 109 | 175 | 2824.0 |
| 4 | 4 | 45 | 2020-12-05 | 117 | 148 | 4060.0 |
| 5 | 5 | 60 | 2020-12-06 | 102 | 127 | 3000.0 |
| 6 | 6 | 60 | 2020-12-07 | 110 | 136 | 3740.0 |
| 7 | 7 | 450 | 2020-12-08 | 104 | 134 | 2533.0 |
| 8 | 8 | 30 | 2020-12-09 | 109 | 133 | 1951.0 |
| 9 | 9 | 60 | 2020-12-10 | 98 | 124 | 2690.0 |
| 10 | 10 | 60 | 2020-12-11 | 103 | 147 | 3293.0 |
| 11 | 11 | 60 | 2020-12-12 | 100 | 120 | 2507.0 |
| 12 | 12 | 60 | 2020-12-12 | 100 | 120 | 2507.0 |
| 13 | 13 | 60 | 2020-12-13 | 106 | 128 | 3453.0 |
| 14 | 14 | 60 | 2020-12-14 | 104 | 132 | 3793.0 |
| 15 | 15 | 60 | 2020-12-15 | 98 | 123 | 2750.0 |
| 16 | 16 | 60 | 2020-12-16 | 98 | 120 | 2152.0 |
| 17 | 17 | 60 | 2020-12-17 | 100 | 120 | 3000.0 |
| 18 | 18 | 45 | 2020-12-18 | 90 | 112 | 0.0 |
| 19 | 19 | 60 | 2020-12-19 | 103 | 123 | 3230.0 |
| 20 | 20 | 45 | 2020-12-20 | 97 | 125 | 0.0 |
| 21 | 1 | 60 | 2020-12-21 | 108 | 131 | 3642.0 |
| 23 | 23 | 60 | 2020-12-23 | 130 | 101 | 3000.0 |
| 24 | 24 | 45 | 2020-12-24 | 105 | 132 | 2460.0 |
| 25 | 25 | 60 | 2020-12-25 | 102 | 126 | 3345.0 |
| 27 | 27 | 60 | 2020-12-27 | 92 | 118 | 2410.0 |
| 28 | 28 | 60 | 2020-12-28 | 103 | 132 | 0.0 |
| 29 | 29 | 60 | 2020-12-29 | 100 | 132 | 2800.0 |
| 30 | 30 | 60 | 2020-12-30 | 102 | 129 | 3803.0 |
| 31 | 31 | 60 | 2020-12-31 | 92 | 115 | 2430.0 |

Endereço do projeto no GITHUB

https://github.com/BrenoSouza2023/Miss-o-Pr-tica-N-vel-3-Mundo-5.git