

RPG0033 - TRATANDO A IMENSIDÃO DOS DADOS

Paulo Wuéliton Horacio Fernandes (202402025919)

507 Polo Tucuruvi – São Paulo/SP

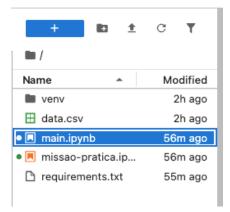
Nível 3: Tratando a Imensidão Dos Dados

Objetivo da Prática

- Descrever como ler um arquivo CSV usando a biblioteca Pandas (Python);
- Descrever como criar um subconjunto de dados a partir de um conjunto existente usando a biblioteca Pandas (Python);
- Descrever como configurar o número máximo de linhas a serem exibidas na visualização de um conjunto de dados usando a biblioteca Pandas (Python);
- Descrever como exibir as primeiras e últimas "N" linhas de um conjunto de dados usando a biblioteca Pandas (Python); Descrever como exibir informações gerais sobre as colunas, linhas e dados de um conjunto de dados usando a biblioteca Pandas (Python);

Microatividade 1: Descrever como ler um arquivo CSV usando a biblioteca Pandas (Python)

 Foi criado e salvo um arquivo chamado data.csv contendo os dados do exercício



2. Foi instalada a biblioteca pandas no projeto e o JupyterLab para execução dos exercícios, as dependências foram adicionadas no projeto usando o comando freeze do pip e salvas no arquivo requirements.txt para que seja possível executar o projeto em outros computadores. Para execução do exercício foi utilizado o comando read_csv para acessar o arquivo data.csv e os dados lidos foram exibidos utilizando o comando print com o to_string do pandas para exibir em modo string.

```
import pandas as pd

data = ''
file = pd.read_csv('./data.csv', sep=';', engine='python', encoding='utf-8')
data = file

print(data.to_string())
```

```
[19]: import pandas as pd
                                                                                                                                                                                                        ⑥↑↓占♀ⅰ
            data = '
            file = pd.read_csv('./data.csv', sep=';', engine='python', encoding='utf-8')
             data = file
            print(data.to_string())
                                                                  Date Pulse Maxpulse Calories
                             60 '2020/12/01'
60 '2020/12/02'
                                                                                              130
                             60 '2020/12/01'
60 '2020/12/02'
60 '2020/12/03'
45 '2020/12/04'
45 '2020/12/06'
60 '2020/12/06'
60 '2020/12/07'
450 '2020/12/08'
30 '2020/12/10'
60 '2020/12/10'
60 '2020/12/11'
60 '2020/12/11'
60 '2020/12/12'
60 '2020/12/12'
60 '2020/12/14'
60 '2020/12/15'
60 '2020/12/16'
60 '2020/12/16'
60 '2020/12/16'
60 '2020/12/16'
60 '2020/12/16'
60 '2020/12/16'
60 '2020/12/16'
60 '2020/12/16'
60 '2020/12/16'
60 '2020/12/21'
45 NaN
60 '2020/12/21'
45 NaN
60 '2020/12/21'
60 '2020/12/21'
60 '2020/12/21'
60 '2020/12/21'
60 '2020/12/21'
60 '2020/12/21'
60 '2020/12/21'
60 '2020/12/21'
60 '2020/12/21'
60 '2020/12/26'
60 '2020/12/28'
60 '2020/12/28'
60 '2020/12/28'
60 '2020/12/30'
60 '2020/12/31'
                                                                              110
117
                                                                                                                       4091
                                                                                                                        3400
                                                                                 109
                                                                                                       175
                                                                                                                       2824
                                                                                 102
                                                                                                      127
                                                                                                                       3000
                                                                                                                        3740
                                                                                 110
                                                                                                       136
                                                                                                       134
                                                                                                                        2533
                                                                                  109
                                                                                                       133
                                                                                                                       1951
            10 10
11 11
                                                                                  103
                                                                                                       147
                                                                                                                       3293
                                                                                                                        2507
                                                                                   100
                                                                                                       120
            12 12
13 13
                                                                                                                        2507
                                                                                   106
                                                                                                       128
                                                                                                                        3453
             14 14
                                                                                                       132
                                                                                                                       3793
            15
16
                                                                                                                       2750
                                                                                                       120
                   16
                                                                                                                       2152
             17 17
                                                                                                                       3000
            18 18
19 19
                                                                                                       112
                                                                                                                        NaN
                                                                                  103
                                                                                                                       3230
            20 20
21 1
                                                                                    97
                                                                                                       125
                                                                                                                    2430 2
                                                                                   108
                                                                                                       131
                                                                                                                       3642
            22 22
23 23
24 24
                                                                                                                        2820
                                                                                  130
                                                                                                      101
                                                                                                                       3000
                                                                                  105
                                                                                                                       2460
                                                                                                       132
            25 25
26 26
27 27
                                                                                                                        3345
                                                                                 100
                                                                                                       120
                                                                                                                        2500
             28 28
29 29
                                                                                 103
                                                                                                       132
                                                                                                                         NaN
                                                                                                                        2800
                                                                                  100
                                                                                                       132
                                 60 '2020/12/30'
60 '2020/12/31'
                                                                                                       129
                                                                                                                        2430
```

Microatividade 2: Descrever como criar um subconjunto de dados a partir de um conjunto existente usando a biblioteca Pandas (Python)

- 1. Criada uma nova variável columns
- Adicionados os dados apenas das 3 primeiras colunas ('ID', 'Duration', 'Date') e exibidos os dados utilizando o comando print:

```
columns = file[['ID', 'Duration', 'Date']]
print(columns.to_string())
```

Microatividade 3: Descrever como configurar o número máximo de linhas a serem exibidas na visualização de um conjunto de dados usando a biblioteca Pandas (Python)

1. Utilizar a configuração max rows para definir o limite de 9999:

```
pd.set_option('display.max_rows', 9999)
print(data.to_string())
```

```
[21]: pd.set_option('display.max_rows', 9999)
                                                                                                                   □↑↓去♀ⅰ
       print(data.to string())
            ID Duration
                                      Date Pulse Maxpulse Calories
                            '2020/12/01'
                                               110
                                                           130
                        60 '2020/12/02'
                                                                     4790
                       60 '2020/12/03'
                                               103
                                                           135
                                                                     3400
                            '2020/12/04'
                                               109
                      45 '2020/12/05'
                                                           148
127
                                               117
                                                                     4060
                      60 '2020/12/06'
                                                                     3000
                                               102
                       60 '2020/12/07'
                      450 '2020/12/08'
                                               104
                                                           134
                                                                     2533
                       30 '2020/12/09'
                                               109
                                                           133
                                                                     1951
                                                                     2690
       9
10
           9
10
                                               98
103
                      60 '2020/12/10'
                                                            124
                      60 '2020/12/11'
                                                           147
                                                                     3293
                           '2020/12/12'
'2020/12/12'
       12
13
            12
                       60 '2020/12/12'
60 '2020/12/13'
60 '2020/12/14'
                                               100
                                                            120
                                                                     2507
                                                                     3453
       14
15
16
17
           14
15
                                               104
                                                           132
                                                                     3793
                       60 '2020/12/15'
                                                            123
                                                                     2750
           16
17
                       60 '2020/12/16'
60 '2020/12/17'
                                               100
                                                           120
                                                                     3000
           18
                            '2020/12/18'
           19
20
1
22
       19
20
                       60 '2020/12/19'
45 '2020/12/20'
                                               103
                                                           123
                                                                     3230
                                                           125
                                                                   2430 2
       21
22
23
                       60 '2020/12/21'
                       45
                                       NaN
                                               100
                                                           119
                                                                     2820
           23
                       60 '2020/12/23'
                                               130
                                                           101
                                                                     3000
           24
25
26
27
28
                                                                     2460
       24
25
26
27
28
                       45 '2020/12/24'
60 '2020/12/25'
                                                            132
                                               102
                                                           126
                                                                     3345
                       20201226
60 '2020/12/27'
60 '2020/12
                                                 92
                                                           118
                                                                     2410
                                               103
                                                           132
                            '2020/12/29'
                                                                     2800
                            '2020/12/30'
                        60
                                                           129
                                                                     3803
                            '2020/12/31'
```

Microatividade 4: Descrever como exibir as primeiras e últimas "N" linhas de um conjunto de dados usando a biblioteca Pandas (Python)

1. Imprimir as 10 primeiras linhas do conjunto de dados utilizando o método head do pandas:

Código:

```
print("Primeiras 10 linhas")
print(file.head(10))
```

Saída:

```
[43]: print("Primeiras 10 linhas")
                                                                                                                       ⑥↑↓占♀ⅰ
       print(file.head(10))
           ID Duration
                                      Date Pulse Maxpulse Calories
                   ration Date
60 '2020/12/01'
60 '2020/12/02'
60 '2020/12/03'
45 '2020/12/04'
45 '2020/12/05'
60 '2020/12/06'
                                               110
                                                            130
                                                103
                                                            135
                                                                      3400
                                                109
                                                            175
                                                                      2824
                                                117
                                                            148
                                                                      4060
                                                102
                                                            127
                                                                      3000
                            '2020/12/07'
                            '2020/12/08'
                      450
                                                104
                                                            134
                                                                      2533
                             '2020/12/09'
                                                109
                                                            133
                                                                      1951
                       30
                       60 '2020/12/10'
```

2. Imprimir as últimas 10 linhas do conjunto de dados utilizando o método tail do pandas:

Código:

```
print("Últimas 10 linhas")
print(file.tail(10))
```

Saída:

```
print("Últimas 10 linhas")
                                                                                                                    □↑↓古♀■
print(file.tail(10))
Últimas 10 linhas
     ID Duration
                                 Date Pulse Maxpulse Calories
          5 NaN
60 '2020/12/23'
45 '2020/12/24'
60 '2020/12/25'
60 '2020/12/25'
60 '2020/12/27'
60 '2020/12/28'
60 '2020/12/29'
22 22
                                         100
                                                       119
23 23
24 24
25 25
                                           130
                                                        101
                                          105
                                                                  2460
                                                        132
                                                        126
26 26
27 27
                                                        120
                                                                  2500
                                            92
                                                        118
                                                                  2410
28 28
29 29
30 30
                                                        132
                                           100
                                                        132
                                                                  2800
                60 '2020/12/30'
                60 '2020/12/31'
                                                                  2430
```

Microatividade 5: Descrever como exibir informações gerais sobre as colunas, linhas e dados de um conjunto de dados usando a biblioteca Pandas (Python)

1. Imprimir as informações gerais do conjunto de dados:

Código:

```
print("Informações gerais")
file.info()
```

Saída:

```
[45]: print("Informações gerais")
                                                                                                                  □↑↓古早■
       file.info()
       Informações 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
            Duration 32 non-null
                                           int64
            Date 31 non-null
Pulse 32 non-null
            Pulse 32 non-null
Maxpulse 32 non-null
Calories 30 non-null
                                           int64
                                           object
       dtypes: int64(4), object(2)
       memory usage: 1.6+ KB
```

2. Imprimir o total de linhas do conjunto de dados:

```
print("Total de linhas")
file.shape[0]
```

```
[46]: print("Total de linhas")
file.shape[0]

Total de linhas
[46]: 32
```

3. Imprimir o total de Colunas do conjunto de dados:

Código:

```
print("Total de colunas")
file.shape[1]
```

Saída:

```
[47]: print("Total de colunas")
file.shape[1]

Total de colunas
[47]: 6
```

4. Imprimir a quantidade de dados nulos:

Código:

```
print("Quantidade de dados nulos")
file.isnull().sum()
```

Saída:

```
•[48]: print("Quantidade de dados nulos")
file.isnul().sum()

Quantidade de dados nulos

[48]: ID 0
Duration 0
Date 1
Pulse 0
Maxpulse 0
Calories 2
dtype: int64
```

5. Imprimir a quantidade de memória utilizada:

```
print("Memória utilizada")
file.memory_usage(deep=True)
```

Saída: атуре: іпточ [7]: print("Memória utilizada") ◎ ↑ ↓ 占 ♀ 🛢 file.memory_usage(deep=True) [7]: Index 256 ID Duration

Missão Prática | Tratando a imensidão dos dados 🗆

Date 1919 Pulse 256 2919 256 25

1656

Maxpulse

Calories

1. Ler o conteúdo do CSV e exibir as informações gerais sobre o conjunto de dados:

Código:

```
import pandas as pd
file = pd.read_csv('./data.csv', sep=';', encoding='utf-8')
print(file.info())
```

Saída:

```
[11]: import pandas as pd
       file = pd.read_csv('./data.csv', sep=';', encoding='utf-8')
[12]: print(file.info())

    ↑ ↓ ∴ ♀ ■
       <class 'pandas.core.frame.DataFrame'>
       RangeIndex: 32 entries, 0 to 31
      object
      4 Maxpulse 32 non-null
5 Calories 30 non-null
dtypes: int64(4), object(2)
                                      object
       memory usage: 1.6+ KB
       None
```

2. Ler as 10 primeiras linhas e as últimas 10 linhas do conjunto de dados:

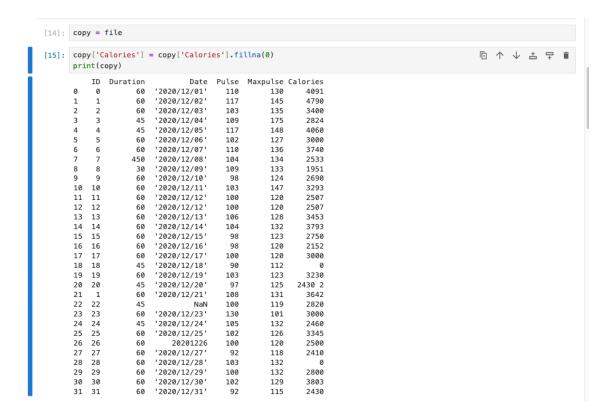
```
print(file.head(10))
print(file.tail(10))
```

```
[13]: print(file.head(10))
                                                                                                                                                                                                    ⑥↑↓≐♀▮
             print(file.tail(10))
                                                               Date Pulse Maxpulse Calories
                                      60 '2020/12/01'
60 '2020/12/02'
                                                                              110
                                                                                                   130
                                                                                                                    4790
                                 60 '2020/12/02'
60 '2020/12/03'
45 '2020/12/04'
45 '2020/12/05'
60 '2020/12/07'
450 '2020/12/08'
30 '2020/12/08'
60 '2020/12/10'
                                                                               117
                                                                                                   145
                                                                              109
117
                                                                                                   175
                                                                                                                    2824
                                                                                                                    4060
                                                                                                   148
                                                                              110
104
                                                                                                    136
                                                                                                                    3740
                                                                                                                    2533
                   7 450 '2020/12/08'
8 30 '2020/12/09'
9 60 '2020/12/10'
ID Duration Date
22 45 NaN
23 60 '2020/12/23'
24 45 '2020/12/24'
25 60 '2020/12/25'
26 60 '2020/12/27'
28 60 '2020/12/27'
28 60 '2020/12/27'
29 60 '2020/12/29'
30 60 '2020/12/30'
31 60 '2020/12/31'
                                                                                                    134
                                                                                                    133
                                                                                 98
                                                                                                    124
                                                                                                                    2690
                                                                           Pulse Maxpulse Calories
100 119 2820
            22
23
24
            22 22
23 23
24 24
25 25
26 26
27 27
                                                                                 130
                                                                                                     101
                                                                                                                     3000
                                                                                                                     2460
                                                                                 105
                                                                                                     132
                                                                                                     126
                                                                                                                     3345
                                                                                100
                                                                                                     120
                                                                                                                     2500
                                                                                                     118
                                                                                                                     2410
            28
29
                                                                                103
100
                                                                                                     132
132
                                                                                                                      NaN
                                                                                                                     2800
                                                                                                     115
                                                                                                                      2430
```

3. Criar uma cópia dos dados em uma nova variável e substituir todos os valores nulos da coluna 'Calories' por 0:

Código:

```
copy = file
copy['Calories'] = copy['Calories'].fillna(0)
print(copy)
```



4. Substituir os valores nulos da coluna 'Date' por '1900/01/01', imprimir o resultado e transformar os dados da coluna 'Date' em datetime utilizando o método 'to_datetime':

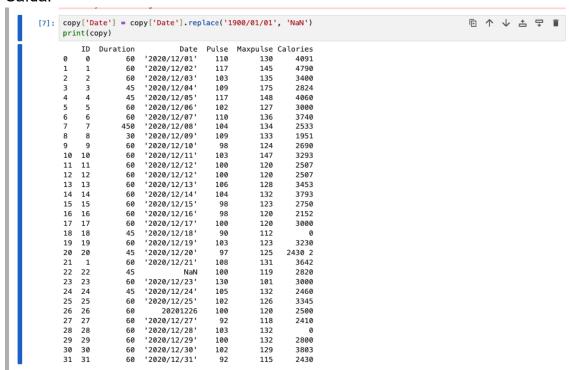
Código:

```
copy['Date'] = copy['Date'].fillna("1900/01/01")
print(copy)
copy['Date'] = pd.to_datetime(copy['Date'])
```

```
•[16]: copy['Date'] = copy['Date'].fillna("1900/01/01")
                                                                                                                                                                                        ⑥↑↓占♀▮
              print(copy)
              copv['Date'] = pd.to datetime(copv['Date'])
                                                                                     Maxpulse Calories
                                              '2020/12/01'
                                       60
                                                                             110
                                                                                                130
                                                                                                               4091
                                              '2020/12/02'
                                              '2020/12/03'
                                        60
                                                                             103
                                                                                                135
                                                                                                               3400
                                               '2020/12/04'
                                        45
                                                                             109
                                                                                                175
                                                                                                               2824
                                        45
                                              '2020/12/05'
                                                                              117
                                                                                                148
                                                                                                               4060
                                              '2020/12/06'
                                        60
                                                                             102
                                                                                                127
                                                                                                               3000
                                              '2020/12/07'
                                      450
                                              '2020/12/08'
                                                                             104
                                                                                                134
                                                                                                               2533
                                               '2020/12/09'
                                                                             109
                                                                                                133
                                                                                                               1951
                      9
                                              '2020/12/10'
                                                                               98
                                                                                                124
                                                                                                               2690
                    10
                                               '2020/12/11'
              10
                                        60
                                                                             103
                                                                                                147
                                                                                                               3293
              12
                     12
                                        60
                                              '2020/12/12'
                                                                             100
                                                                                                120
                                                                                                               2507
                                               '2020/12/13'
                                                                                                               3453
              13
                     13
                                        60
                                                                             106
                                                                                                128
              14
                     14
                                              '2020/12/14'
                                                                             104
                                                                                                132
                                                                                                               3793
                                              '2020/12/15'
              15
                     15
                                        60
                                                                               98
                                                                                                123
                                                                                                               2750
              17
                     17
                                        60
                                              '2020/12/17'
                                                                             100
                                                                                                120
                                                                                                               3000
                                               '2020/12/18'
              18
                    18
                                                                               90
                                                                                                112
              19
                     19
                                              '2020/12/19'
                                                                             103
                                                                                                123
                                                                                                               3230
                                               '2020/12/20'
              20
                    20
                                        45
                                                                               97
                                                                                                125
                                                                                                            2430 2
                                              '2020/12/21'
              22
                    22
                                        45
                                                  1900/01/01
                                                                             100
                                                                                                119
                                                                                                               2820
              23
                                              '2020/12/23'
                     23
                                        60
                                                                             130
                                                                                                101
                                                                                                               3000
              24
25
                    24
25
                                        45
                                              '2020/12/24'
                                                                             105
                                                                                                132
                                                                                                               2460
                                              '2020/12/25'
                                        60
                                                                             102
                                                                                                126
                                                                                                               3345
                                              '2020/12/27'
              27
                    27
                                        60
                                                                               92
                                                                                                118
                                                                                                               2410
              28
                    28
                                               '2020/12/28'
                                                                             103
                                                                                                132
              29
                    29
                                        60
                                              '2020/12/29'
                                                                             100
                                                                                                132
                                                                                                               2800
                                               '2020/12/30'
              30
                     30
                                        60
                                                                             102
                                                                                                129
                                                                                                               3803
                                              '2020/12/31'
                                                                                                               2430
              \textbf{File } \sim / \texttt{Documents/python/missao-pratica-nivel3-mundo5/venv/lib/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/pandas/core/tools/datetim/python3.13/site-packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages/packages
              es.py:1067, in to_datetime(arg, errors, dayfirst, yearfirst, utc, format, exact, unit, infer_datetime_format, or
                   1065
                                            result = arg.map(cache_array)
                                    else:
                                           values = convert_listlike(arg._values, format)
result = arg._constructor(values, index=arg.index, name=arg.name)
              -> 1067
                   1069 elif isinstance(arg, (ABCDataFrame, abc.MutableMapping)):
              File ~/Documents/python/missao-pratica-nivel3-mundo5/venv/lib/python3.13/site-packages/pandas/core/tools/datetim
              es.py:433, in _convert_listlike_datetimes(arg, format, name, utc, unit, errors, dayfirst, yearfirst, exact)
431 # `format` could be inferred, or user didn't ask for mixed-format parsing.
                     432 if format is not None and format != "mixed":
433    return _array_strptime_with_fallback(arg, name, utc, format, exact, errors)
435    result, tz_parsed = objects_to_datetime64(
               --> 433
                     436
                     437
                                    dayfirst=dayfirst,
                                                 allow_object=True,
                     442 )
                     444 if tz_parsed is not None:
                                  # We can take a shortcut since the datetime64 numpy array
# is in UTC
                     445
              File ~/Documents/python/missao-pratica-nivel3-mundo5/venv/lib/python3.13/site-packages/pandas/core/tools/datetim
              es.py:467, in _array_strptime_with_fallback(arg, name, utc, fmt, exact, errors)
                      456 def _array_strptime_with_fallback(
                             arg,
                     458
                                   462
                                                  errors: str,
                     464
                                    Call array_strptime, with fallback behavior depending on 'errors'.
                     466
                                    result, tz_out = array_strptime(arg, fmt, exact=exact, errors=errors, utc=utc)
               --> 467
                     469
                                           unit = np.datetime_data(result.dtype)[0]
              File strptime.pyx:501, in pandas._libs.tslibs.strptime.array_strptime()
              File strptime.pyx:451, in pandas._libs.tslibs.strptime.array_strptime()
              File strptime.pyx:583, in pandas._libs.tslibs.strptime._parse_with_format()
              ValueError: time data "1900/01/01" doesn't match format "'%Y/%m/%d'", at position 22. You might want to try:
                     - passing `format` if your strings have a consistent format;
- passing `format='ISO8601'` if your strings are all ISO8601 but not necessarily in exactly the same format;
- passing `format='mixed'`, and the format will be inferred for each element individually. You might want to
              use `dayfirst` alongside this.
```

5. Substituir o valor '1900/01/01' por 'NaN' da coluna 'Date'

```
copy['Date'] = copy['Date'].replace('1900/01/01', 'NaN')
print(copy)
```



6. Transformar os dados da coluna 'Date' para datetime usando o método 'to_datetime':

Código:

```
copy['Date'] = pd.to_datetime(copy['Date'])
```

```
File ~/Documents/python/missao-pratica-nivel3-mundo5/venv/lib/python3.13/site-packages/pandas/core/tools/datetim
 es.py:1067, in to_datetime(arg, errors, dayfirst, yearfirst, utc, format, exact, unit, infer_datetime_format, or
igin, cache)
1065
                     result = arg.map(cache array)
   1066
              values = convert_listlike(arg._values, format)
result = arg._constructor(values, index=arg.index, name=arg.name)
 -> 1067
   1069 elif isinstance(arg, (ABCDataFrame, abc.MutableMapping)):
File ~/Documents/python/missao-pratica-nivel3-mundo5/venv/lib/python3.13/site-packages/pandas/core/tools/datetim
es.py:433, in _convert_listlike_datetimes(arg, format, name, utc, unit, errors, dayfirst, yearfirst, exact) 431 # `format` could be inferred, or user didn't ask for mixed-format parsing.
   432 if format is not None and format != "mixed":
> 433 return _array_strptime_with_fallback(arg, name, utc, format, exact, errors)
435 result, tz_parsed = objects_to_datetime64(
     436 arg,
437 dayfirst=dayfirst,
(...) 441 allow_object=True,
     444 if tz_parsed is not None:
     # We can take a shortcut since the datetime64 numpy array
446 # is in UTC
File ~/Documents/pvthon/missao-pratica-nivel3-mundo5/venv/lib/python3.13/site-packages/pandas/core/tools/datetim
es.py:467, in _array_strptime_with_fallback(arg, name, utc, fmt, exact, errors)
      456 def _array_strptime_with_fallback(
     457 arg,
458 name,
(...) 462
     457
                           errors: str,
      463 ) -> Index:
     464
                Call array_strptime, with fallback behavior depending on 'errors'.
     465
           result, tz_out = array_strptime(arg, fmt, exact=exact, errors=errors, utc=utc)
if tz_out is not None:
 --> 467
                     unit = np.datetime_data(result.dtype)[0]
File strptime.pyx:501, in pandas._libs.tslibs.strptime.array_strptime()
File strptime.pvx:451, in pandas. libs.tslibs.strptime.array strptime()
File strptime.pyx:583, in pandas._libs.tslibs.strptime._parse_with_format()
ValueError: time data "20201226" doesn't match format "'%Y/%m/%d'", at position 26. You might want to try:

- passing `format` if your strings have a consistent format;

- passing `format='IS08601'` if your strings are all IS08601 but not necessarily in exactly the same format;

- passing `format='mixed'`, and the format will be inferred for each element individually. You might want to use `dayfirst` alongside this.
```

7. Transformar o valor '20201226' em uma data válida usando o 'replace' e o 'to_datetime' e converter todos os dados da coluna 'Date' para datas:

Código:

```
copy['Date'] = copy['Date'].replace('20201226', pd.to_datetime('2020/12/26',
format='%Y/%m/%d'))
copy['Date'] = pd.to_datetime(copy['Date'])
print(copy)
```

```
copy['Date'] = copy['Date'].replace('20201226', pd.to\_datetime('2020/12/26', format='%Y/%m/%d'))
copy['Date'] = pd.to_datetime(copy['Date'])
print(copy)
         Duration
                          Date
                                        Maxpulse Calories
                60 2020-12-01
                                   110
                                              130
                                                       4091
                60 2020-12-02
                                   117
                                              145
                60 2020-12-03
                                   103
                                              135
                                                       3400
                45 2020-12-04
                                   109
                                              175
                                                       2824
                45 2020-12-05
                                   117
                60 2020-12-06
                                   102
                                              127
                                                       3000
                60 2020-12-07
                                                       3740
                                   110
                                              136
               450 2020-12-08
                                              134
                                                       2533
8
                30 2020-12-09
                                   109
                                              133
                                                       1951
                60 2020-12-10
                                    98
                                                       2690
10
11
    10
11
                                              147
120
                60 2020-12-11
                                   103
                                                       3293
                60 2020-12-12
                                                       2507
                                   100
12
                60 2020-12-12
                                   100
                                                       2507
    13
14
13
14
15
16
17
                60 2020-12-13
                                   106
                                              128
                                                       3453
                60 2020-12-14
                                              132
                                                       3793
                                   104
    15
                60 2020-12-15
                                    98
                                              123
                                                       2750
    16
                60 2020-12-16
                                    98
                                              120
                                                       2152
    17
                60 2020-12-17
                                              120
18
19
    18
19
                45 2020-12-18
                                    90
                                              112
                60 2020-12-19
                                                       3230
                                   103
                                              123
20
21
22
23
24
    20
                45 2020-12-20
                                    97
                                              125
                                                     2430 2
                60 2020-12-21
                                   108
                                              131
                                                       3642
    22
                                   100
                                                       2820
    23
24
                60 2020-12-23
                                   130
                                              101
                                                       3000
                45 2020-12-24
                                                       2460
                                   105
                                              132
25
26
27
    25
                60 2020-12-25
                                   102
                                              126
                                                       3345
    26
                60 2020-12-26
                                   100
                                              120
                                                       2500
    27
                60 2020-12-27
                                    92
                                              118
                                                       2410
28
29
    28
29
                60 2020-12-28
                                   103
                                              132
                                                       2800
                60 2020-12-29
                                   100
                                              132
                60 2020-12-30
31
    31
                60 2020-12-31
                                    92
                                              115
                                                       2430
```

8. Remover os registros nulos:

Código:

```
copy = copy.dropna(subset=['Date'])
print(copy)
```

```
[10]: copy = copy.dropna(subset=['Date'])
                                                                                                           回个小占
                                                                                                                           무
                                                                                                                               Ĥ
 [11]: print(copy)
             ID
                Duration
                                 Date
                                        Pulse Maxpulse Calories
                       60 2020-12-01
                                          110
                                                     130
                                                              4091
                        60 2020-12-02
                                                     145
                                                              4790
                        60 2020-12-03
                                          103
                                                     135
                                                              3400
                        45 2020-12-04
                                          109
                                                     175
                                                              2824
                        45 2020-12-05
                                          117
                                                     148
                                                              4060
                        60 2020-12-06
                                                     127
                                                              3000
                                          102
                        60 2020-12-07
                                          110
                                                     136
                       450 2020-12-08
                                          104
                                                     134
                                                              2533
                       30 2020-12-09
                                                     133
                                                              1951
                                          109
                        60 2020-12-10
                                                              2690
        10
11
             10
                        60 2020-12-11
                                          103
                                                     147
                                                              3293
             11
                        60 2020-12-12
                                                     120
                                                              2507
                                          100
            12
13
        12
13
14
15
16
                        60 2020-12-12
                                          100
                                                     120
                                                              2507
                        60 2020-12-13
                                          106
                                                     128
                                                              3453
                        60 2020-12-14
                                                     132
                                                              3793
            15
16
                        60 2020-12-15
                                           98
                                                     123
                                                              2750
                        60 2020-12-16
                                                     120
                                           98
                                                              2152
        17
18
             17
                        60 2020-12-17
                                          100
                                                     120
                                                              3000
             18
                        45 2020-12-18
                                           90
                                                     112
        19
                        60 2020-12-19
                                                              3230
                                          103
                                                     123
        20
21
             20
                        45 2020-12-20
                                           97
                                                     125
                                                           2430 2
                        60 2020-12-21
                                                              3642
                                          108
                                                     131
        23
24
25
             23
                        60 2020-12-23
                                          130
                                                              3000
             24
                        45 2020-12-24
                                          105
                                                     132
                                                              2460
             25
                        60 2020-12-25
                                          102
                                                     126
                                                              3345
        26
27
            26
27
                        60 2020-12-26
                                          100
                                                     120
                                                              2500
                        60 2020-12-27
                                           92
                                                              2410
                                                     118
        28
                        60 2020-12-28
                                                     132
        29
30
                                                              2800
            29
                        60 2020-12-29
                                          100
                                                     132
             30
                        60 2020-12-30
                                          102
                                                     129
                                                              3803
                        60 2020-12-31
                                                              2430
```