UNIVERSIDADE ESTACIO DE SÁ



POLO PARANGABA – FORTALEZA/CE

DESENVOLVIMENTO FULL STACK - 22.3

Relatório da Missão Prática | Nível 3 | Mundo 5

Aluno:	Jefferson Ponte Pessoa
Professor:	Altamira de Souza Queiroz
Repositório:	https://github.com/jeffersonkako/N3M5

Título da Prática:

TRATANDO A IMENSIDÃO DOS DADOS

Objetivo da Pratica:

- 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);

Códigos:

memory usage: 1.6+ KB

```
In [1]: import pandas as pd
          data = pd.read_csv('data.csv', sep=';', engine='python', encoding='utf-8')
In [3]:
          print("Informações gerais sobre o conjunto de dados:")
          display(data.info())
        Informações gerais sobre o conjunto de dados:
        <class 'pandas.core.frame.DataFrame'>
RangeIndex: 32 entries, 0 to 31
       Data columns (total 6 columns):
                        Non-Null Count Dtype
            Column
             ID
                                          int64
                        32 non-null
             Duration 32 non-null
             Date
                        31 non-null
                                          object
             Pulse
                        32 non-null
                                          int64
             Maxpulse 32 non-null
Calories 30 non-null
                                          int64
                                          float64
        dtypes: float64(1), int64(4), object(1)
```

```
In [4]:
          print("Primeiras 5 linhas:")
          display(data.head())
        Primeiras 5 linhas:
           ID Duration
                                Date Pulse Maxpulse Calories
        0 0
                     60 '2020/12/01'
                                                   130
                                                          4091.0
                                        110
In [7]:
           data_copia['Calories'] = data_copia['Calories'].fillna(0)
          print("Dados após substituir nulos em 'Calories':")
display(data_copia)
        Dados após substituir nulos em 'Calories':
             ID Duration
                            Date Pulse Maxpulse Calories
         0
             0
                       60 '2020/12/01'
                                                            4091.0
                                           110
                                                     130
         1
                       60 '2020/12/02'
                                           117
                                                      145
                                                            4790.0
         2
                       60
                          '2020/12/03'
                                           103
                                                      135
                                                            3400.0
         3
                       45 '2020/12/04'
                                                            2824.0
                                           109
                                                      175
         4
              4
                           '2020/12/05'
                                                      148
                                                            4060.0
         5
                       60 '2020/12/06'
                                          102
                                                     127
                                                            3000.0
             5
         6
                           '2020/12/07'
                                                            3740.0
         7
                                                            2533.0
                      450 '2020/12/08'
                                           104
                                                      134
         8
                       30 '2020/12/09'
                                           109
                                                      133
                                                             1951.0
         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
                                                            2507.0
                       60 '2020/12/12'
                                           100
                                                      120
        13
                       60
                           '2020/12/13'
                                                     128
                                                            3453.0
            13
                                           106
        14
                                                            3793.0
            14
                       60 '2020/12/14'
                                           104
                                                      132
        15
            15
                           '2020/12/15'
                                                      123
                                                            2750.0
        16
                       60 '2020/12/16'
                                                            2152.0
            16
                                           98
                                                      120
         17
            17
                       60 '2020/12/17'
                                           100
                                                      120
                                                            3000.0
        17 17
                       60 '2020/12/17'
                                           100
                                                      120
                                                            3000.0
        18
            18
                       45 '2020/12/18'
                                                                0.0
                                           90
                                                      112
        19
            19
                       60
                           '2020/12/19'
                                           103
                                                      123
                                                            3230.0
        20 20
                          '2020/12/20'
                                           97
                                                            2430.0
                       45
                                                      125
        21 21
                       60
                          '2020/12/21'
                                           108
                                                      131
                                                            3642.0
        22 22
                       45
                                  NaN
                                           100
                                                      119
                                                            2820.0
        23 23
                       60 '2020/12/23'
                                          130
                                                      101
                                                            3000.0
        24 24
                       45 '2020/12/24'
                                                             2460.0
        25 25
                       60 '2020/12/25'
                                           102
                                                            3345.0
                                                      126
        26 26
                             20201226
                                           100
                                                      120
                                                             2500.0
        27 27
                                                             2410.0
                       60 '2020/12/27'
                                           92
                                                      118
        28 28
                       60 '2020/12/28'
                                                      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'
                                                            2430.0
                                          92
                                                      115
  In [8]: data_copia['Date'] = data_copia['Date'].fillna('1900/01/01')
    print("Dados após substituir nulos em 'Date':")
    display(data_copia)
        Dados após substituir nulos em 'Date':
                          Date Pulse Maxpulse Calories
           0
                  60 '2020/12/01'
                                             4091.0
                                110
                                        130
        1 1
                  60 '2020/12/02' 117
                                       145 4790.0
         2 2
                  60 '2020/12/03'
                                103
                                        135
                                             3400.0
                  45 '2020/12/04'
                                        175 2824.0
                   45 '2020/12/05'
                  60 '2020/12/06' 102
         5 5
                                        127 3000.0
         6
           6
                  60 '2020/12/07'
                                110
                                        136
                                             3740.0
        7 7
                 450 '2020/12/08' 104
                                        134 2533.0
                  30 '2020/12/09'
                                         133
                                              1951.0
                  60 '2020/12/10'
                                        124
```

147

120 2507.0

120 2152.0

103

100

106

3293.0

120 2507.0

128 3453.0

123 2750.0

10 10

12 12

13 13

15 15

16 16

11 11

60 '2020/12/11'

60 '2020/12/12'

60 '2020/12/16'

60 '2020/12/13'

'2020/12/14'

60 '2020/12/15' 98

17 17 60 '2020/12/17' 100 120 3000.0

60 '2020/12/12' 100

```
In [9]:
         data_copia['Date'] = pd.to_datetime(data_copia['Date'], errors='coerce')
print("Dados após a conversão inicial para datetime:")
          display(data_copia)
       Dados após a conversão inicial para datetime:
            ID Duration
                                Date Pulse Maxpulse Calories
             0
                      60 2020-12-01
                                        110
                                                   130
                                                         4091.0
         0
                      60 2020-12-02
                                                   145
                                                         4790.0
         2
                      60 2020-12-03
                                                         3400.0
                                        103
                                                   135
         3
                      45
                         2020-12-04
                                        109
                                                   175
                                                         2824.0
                      45 2020-12-05
                                        117
                                                   148
                                                         4060.0
         5
                     60 2020-12-06
                                                         3000.0
             5
                                        102
                                                   127
         6
                      60 2020-12-07
                                                         3740.0
         7
                    450 2020-12-08
                                                         2533.0
                                        104
                                                   134
         8
             8
                      30
                         2020-12-09
                                        109
                                                   133
                                                          1951.0
        9
                                                         2690.0
             9
                     60
                          2020-12-10
                                        98
                                                   124
        10
            10
                      60
                          2020-12-11
                                        103
                                                   147
                                                         3293.0
        11
                         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
                                                         3793.0
                                                   132
                     60 2020 12 15
                                       00
                                                  122
                                                         2750.0
            data_copia['Date'] = data_copia['Date'].replace('1900-01-01', pd.NaT)
print("Dados após substituir '1900/01/01' por NaN:")
display(data_copia)
          Dados após substituir '1900/01/01' por NaN:
              ID Duration
                                  Date Pulse Maxpulse Calories
           0
               0
                       60 2020-12-01
                                          110
                                                           4091.0
                                                    130
           1
                       60 2020-12-02
                                          117
                                                    145
                                                           4790.0
           2
                       60 2020-12-03
                                                    135
                                                           3400.0
                       45 2020-12-04
                                                    175
           3
               3
                                         109
                                                           2824.0
           4
               4
                       45 2020-12-05
                                          117
                                                    148
                                                           4060.0
                       60 2020-12-06
                                         102
                                                    127
                                                           3000.0
           6
               6
                       60 2020-12-07
                                                           3740.0
                                          110
                                                    136
           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
                            2020-12-11
                                          103
                                                    147
                                                           3293.0
                                                           2507.0
          11
               11
                       60 2020-12-12
                                         100
                                                    120
          12 12
                       60 2020-12-12
                                         100
                                                    120
                                                           2507.0
          13 13
                       60 2020-12-13 106
                                               128
                                                           3453.0
In [12]:
           data_copia['Date'] = data_copia['Date'].replace('20201226', '2020/12/26')
In [13]:
           data_copia['Date'] = pd.to_datetime(data_copia['Date'], errors='coerce')
           print("Dados após corrigir '20201226' e converter para datetime:")
display(data_copia)
         Dados após corrigir '20201226' e converter para datetime:
              ID Duration
                                 Date Pulse Maxpulse Calories
              0
                       60 2020-12-01
                                         110
                                                    130
                                                           4091.0
                       60 2020-12-02
                                         117
                                                    145
                                                          4790.0
          2
              2
                       60 2020-12-03
                                         103
                                                    135
                                                          3400.0
          3
                       45 2020-12-04
                                         109
                                                    175
          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
                                                           3740.0
                                         110
                                                    136
          7
              7
                      450 2020-12-08
                                         104
                                                    134
                                                           2533.0
          8
              8
                       30 2020-12-09
                                                    133
                                                           1951.0
          9
                       60 2020-12-10
              9
                                          98
                                                    124
                                                          2690.0
         10
             10
                       60 2020-12-11
                                         103
                                                    147
                                                           3293.0
             11
                       60 2020-12-12
         11
                                         100
                                                    120
                                                           2507.0
         12
             12
                       60 2020-12-12
                                         100
                                                    120
                                                           2507.0
```

13

14

15 15

14

2020-12-13

2020-12-15

60 2020-12-14

60

3453.0

3793.0

2750.0

128

132

123

104

In [14]: data_copia.dropna(subset=['Date'], inplace=True) In [15]: print("Dados finais após todas as transformações e remoção de nulos:") display(data_copia) Dados finais após todas as transformações e remoção de nulos: ID Duration Date Pulse Maxpulse Calories 60 2020-12-01 110 130 4091.0 **1** 1 60 2020-12-02 117 145 4790.0 60 2020-12-03 103 135 3400.0 **3** 3 45 2020-12-04 109 175 2824.0 45 2020-12-05 117 148 4060.0 **5** 5 60 2020-12-06 102 127 3000.0 60 2020-12-07 110 136 3740.0 **7** 7 450 2020-12-08 104 134 2533.0 30 2020-12-09 109 133 1951.0 60 2020-12-10 98 124 2690.0 60 2020-12-11 103 147 3293.0 **11** 11 60 2020-12-12 100 120 2507.0 60 2020-12-12 100 120 2507.0 60 2020-12-13 106 128 3453.0 60 2020-12-14 104 132 3793.0 **15** 15 60 2020-12-15 98 123 2750.0

Código completo no repositorio:

https://github.com/jeffersonkako/N3M5

Ferramentas:

VSCODE PERFIL PYTHON PADRÃO DO VSCODE PYTHON3 JUPYTER