



## Faculdade Estácio - Campus - Monte Castelo

**Curso:** Desenvolvimento Full Stack

**Disciplina:** Tratando a Imensidão dos Dados

**Número da Turma:** RPG0033

**Semestre Letivo:** 5

**Integrante:** Lucas de Oliveira dos Santos

### Missão Prática: ***Tratando a Imensidão dos Dados***

Esta Missão Prática tem como objetivo aplicar os conhecimentos em manipulação e tratamento de dados utilizando a biblioteca Pandas em Python. Este documento reúne todos os scripts e resultados produzidos, organizando os de forma a demonstrar as técnicas aplicadas para a limpeza e análise de um conjunto de dados real.

O objetivo desta missão prática foi realizar a limpeza e a preparação de um conjunto de dados para futuras análises. O conjunto de dados foi tratado para corrigir valores nulos, ajustar formatos de data, e corrigir valores incorretos, como IDs e durações. Também foram removidas duplicidades e linhas com valores indesejados.

### **Passos Realizados:**

#### **1. Leitura do Arquivo CSV:**

O arquivo foi carregado e as primeiras e últimas linhas foram exibidas.

### Dados Antes do Tratamento:

ID;Duration;Date;Pulse;Maxpulse;Calories

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

### 2. Substituição de Valores Nulos:

Valores nulos na coluna **Calories** foram substituídos por **0**.

Datas ausentes na coluna **Date** foram preenchidas com a data padrão 01/01/2020 .

### 3. Correção de Datas:

Datas no formato **YYYYMMDD** foram convertidas para **YYYY/MM/DD** .

A coluna **Date** foi convertida para o tipo **datetime** . 4.

### Correção de Duração e ID:

A duração incorreta de 450 na linha 7 foi corrigida para 45.

O ID incorreto na linha 21 foi ajustado de 1 para 21.

### 5. Correção de Valores na Coluna **Calories** :

Na linha 20, o valor "2430 2" foi corrigido para "2430".

### 6. Remoção de Linhas Duplicadas e Zeradas:

Linhas duplicadas foram removidas.

Linhas onde a coluna **Calories** tinha valor 0 foram excluídas.

## Resultados Finais:

Após todas as transformações, o DataFrame final foi limpo e preparado, resultando em um conjunto de dados consistente e pronto para análise.

### Dados Após o Tratamento:

ID	Duration	Date	Pulse	Maxpulse	Calories
4091	1	2020-12-01	110	130	130
3400	3	2020-12-02	103	135	135
4060	5	2020-12-03	103	135	135
3740	7	2020-12-04	109	148	148
1951	9	2020-12-05	117	148	148
3293	11	2020-12-06	102	136	136
2507	13	2020-12-07	110	136	136
		2020-12-08	104	133	133
		2020-12-09	109	133	133
		2020-12-10	98	124	147
		2020-12-11	100	120	120
		2020-12-12	100	120	120
		2020-12-13	106	128	132
		2020-12-14	104	132	132

3793 | | 15 | 60 | 2020-12-15 | 98 | 123 | 2750 | | 16 | 60 | 2020-12-16 | 98 | 120 |  
2152 | | 17 | 60 | 2020-12-17 | 100 | 120 | 3000 | | 18 | 45 | 2020-12-18 | 90 | 112 | 0  
	19	60	2020-12-19	103	123	3230		20	45	2020-12-20	97	125	2430
21	60	2020-12-21	108	131	3642		23	60	2020-12-23	130	101	3000	
24	45	2020-12-24	105	132	2460		25	60	2020-12-25	102	126	3345	
26	60	2020-12-26	100	120	2500		27	60	2020-12-27	92	118	2410	
28 | 60 | 2020-12-28 | 103 | 132 | 0 | | 29 | 60 | 2020-12-29 | 100 | 132 | 2800 | | 30 |  
60 | 2020-12-30 | 102 | 129 | 3803 | | 31 | 60 | 2020-12-31 | 92 | 115 | 2430 |

## Conclusão:

A aplicação das técnicas de manipulação e limpeza de dados com a biblioteca Pandas foi essencial para transformar um conjunto de dados bruto em uma base de dados confiável. Este processo assegurou que os dados estivessem em um formato adequado para análises futuras, permitindo a extração de insights precisos e significativos.