

Tratamento de Dados do Dataset de Hipertensão

Seu Nome

31 de outubro de 2025

Objetivo

O objetivo desta apresentação é mostrar o pré-processamento do dataset de hipertensão e hábitos de atividade física, com foco no tratamento de dados faltantes e criação de variáveis derivadas.

Consumo Semanal de Frutas e Verduras

P018 P019 consumo_semanal_frutas			
0	3.0	NaN	3
1	1.0	NaN	1
2	3.0	NaN	3
3	5.0	1.0	5
4	4.0	NaN	4
5	7.0	2.0	14
6	2.0	NaN	2
7	7.0	1.0	7
8	7.0	1.0	7
9	0.0	NaN	0

P00901 P01001 consumo_semanal_verdura_legume			
0	1.0	NaN	0
1	0.0	NaN	0
2	5.0	2	10
3	3.0	NaN	0
4	0.0	NaN	0
5	7.0	1	7
6	4.0	NaN	0
7	7.0	1	7
8	7.0	1	7
9	0.0	NaN	0

Doenças Cardíacas e Esporte Praticado

	Q08306	Q08307	Q08308	Q08309	Q08310	Q08311	infarto	angina	insuficiencia_cardiaca	arritmia	outra_doenca_cardiaca	
0	2	NaN	NaN	NaN	NaN	NaN	2	2	2	2	2	2
1	2	NaN	NaN	NaN	NaN	NaN	2	2	2	2	2	2
2	2	NaN	NaN	NaN	NaN	NaN	2	2	2	2	2	2
3	2	NaN	NaN	NaN	NaN	NaN	2	2	2	2	2	2
4	2	NaN	NaN	NaN	NaN	NaN	2	2	2	2	2	2
5	2	NaN	NaN	NaN	NaN	NaN	2	2	2	2	2	2
6	2	NaN	NaN	NaN	NaN	NaN	2	2	2	2	2	2
7	1	1	2	1	2	2	3	2	1	2	2	2
8	2	NaN	NaN	NaN	NaN	NaN	2	2	2	2	2	2
9	1	2	2	2	1	2	2	2	2	1	2	2

	P034	P036	esporte_praticado	
0	1	1		1
1	2	NaN		0
2	2	NaN		0
3	1	1		1
4	2	NaN		0
5	1	6		6
6	1	9		9
7	1	1		1
8	1	5		5
9	2	NaN		0

Horas e Minutos de Atividade Física

	P034	P03701	horas_atv_fis
0	1	0.0	0
1	2	NaN	0
2	2	NaN	0
3	1	0.0	0
4	2	NaN	0
5	1	1.0	1
6	1	1.0	1
7	1	1.0	1
8	1	1.0	1
9	2	NaN	0

	P034	P03702	minutos_atv_fis
0	1	30.0	30
1	2	NaN	0
2	2	NaN	0
3	1	30.0	30
4	2	NaN	0
5	1	0.0	0
6	1	0.0	0
7	1	0.0	0
8	1	0.0	0
9	2	NaN	0

Idade de Doença Cardíaca e Medicação

Q06306 Q064 idade_doenca_cardiaca			
0	2	NaN	99
1	2	NaN	99
2	2	NaN	99
3	2	NaN	99
4	2	NaN	99
5	2	NaN	99
6	2	NaN	99
7	1	15.0	15
8	2	NaN	99
9	1	56.0	56

Q00503 Q00601 Q00602 medicacao_hipertensao Q00602_novo					
0	1	1	NaN	1	0
1	1	2	6	2	6
2	1	1	NaN	1	0
3	1	1	NaN	1	0
4	1	3	7	3	7
5	1	1	NaN	1	0
6	1	1	NaN	1	0
7	1	1	NaN	1	0
8	1	1	NaN	1	0
9	1	1	NaN	1	0

Tratamientos P033

P027	P03301	P03301_tratado		P027	P03302	P03302_tratado	
0	2	1	1	0	2	1	1
1	1	NaN	0	1	1	NaN	0
2	1	NaN	0	2	1	NaN	0
3	1	NaN	0	3	1	NaN	0
4	1	NaN	0	4	1	NaN	0
5	1	NaN	0	5	1	NaN	0
6	2	1	1	6	2	1	1
7	1	NaN	0	7	1	NaN	0
8	3	1	1	8	3	1	1
9	1	NaN	0	9	1	NaN	0

Mais Tratamentos P033 e Prática Semanal

P027 P03303 P03303_tratado				P034 P035 pratica_semanal_atv_fis			
0	2	1	1	0	1	1.0	1
1	1	NaN	0	1	2	NaN	0
2	1	NaN	0	2	2	NaN	0
3	1	NaN	0	3	1	2.0	2
4	1	NaN	0	4	2	NaN	0
5	1	NaN	0	5	1	6.0	6
6	2	1	1	6	1	2.0	2
7	1	NaN	0	7	1	2.0	2
8	3	1	1	8	1	5.0	5
9	1	NaN	0	9	2	NaN	0

Novas Variáveis Alimentares

Q00401 Q00502 Q00502_novo				P023 P02401 tipo_leite_consumido			
0	1	NaN	0	0	7.0	2	2
1	2	4	4	1	7.0	2	2
2	2	4	4	2	7.0	2	2
3	2	2	2	3	7.0	1	1
4	2	4	4	4	2.0	2	2
5	1	NaN	0	5	0.0	NaN	0
6	3	10	10	6	7.0	3	3
7	1	NaN	0	7	0.0	NaN	0
8	3	10	10	8	2.0	2	2
9	1	NaN	0	9	7.0	2	2

Tipos de Bebidas Consumidas

P02002 P02102 tipo_refri_consumido				P02001 P02101 tipo_suco_consumido			
0	1.0	2	2	0	7.0	2	2
1	0.0	NaN	0	1	0.0	NaN	0
2	0.0	NaN	0	2	3.0	2	2
3	0.0	NaN	0	3	0.0	NaN	0
4	1.0	2	2	4	1.0	2	2
5	0.0	NaN	0	5	1.0	2	2
6	0.0	NaN	0	6	0.0	NaN	0
7	7.0	2	2	7	0.0	NaN	0
8	1.0	3	3	8	0.0	NaN	0
9	0.0	NaN	0	9	4.0	2	2

Remoção de Linhas com Valores Nulos

Após a análise, decidiu-se remover todas as linhas que possuíam valores nulos. No total, foram removidas **1484 linhas**, preservando as colunas mais completas do dataset.

	Q028	Q00401	Q00503	Q003	horas_atv_fis	minutos_atv_fis	esporte_praticado
21	NaN	NaN	NaN	NaN	0	0	0
35	NaN	NaN	NaN	NaN	0	0	0
54	NaN	NaN	NaN	NaN	0	0	0
58	NaN	NaN	NaN	NaN	1	30	5
71	NaN	NaN	NaN	NaN	1	0	1
73	1	2	1	47.0	<NA>	<NA>	NaN
76	NaN	NaN	NaN	NaN	2	0	17
82	NaN	NaN	NaN	NaN	0	0	0
85	NaN	NaN	NaN	NaN	0	0	0
111	1	1	1	22.0	<NA>	<NA>	NaN

Figura: Visualização das colunas que possuíam dados nulos antes da remoção.

Conclusão

- ▶ Foi realizada a análise e visualização de valores nulos no dataset.
- ▶ Criadas transformações e métricas derivadas para monitorar a completude dos dados.
- ▶ Linhas com dados faltantes foram removidas (1484 linhas) para garantir integridade do dataset.