



CARLOS DANIEL DE MOURA SANTOS - 31811BSI022
ELTON LUNARDI -32011BSI009
NICK BRUNO YUDI ISHIDA - 32011BSI011
PABLO SOARES DIAS - 31821BSI007

MINERAÇÃO DE DADOS
TRABALHO I

MONTE CARMELO
2022

CARLOS DANIEL DE MOURA SANTOS - 31811BSI022
ELTON LUNARDI -32011BSI009
NICK BRUNO YUDI ISHIDA - 32011BSI011
PABLO SOARES DIAS - 31821BSI007

GS1556 - MINERAÇÃO DE DADOS
TRABALHO I

Trabalho de Mineração de dados.
Este trabalho tem por finalidade
relatório relatando os problemas,
descrição das bases, tarefas executadas.
Prof.: Carlos Cesar Mansur Tuma.

Relatório do Trabalho

Encontrar conhecimento relacionando idade com consumo de produtos alimentares.

No.	1: código Numeric	2: nome Nominal	3: idade Numeric	4: endereço Nominal
1	1.0	carlos	34.0	x32
2	2.0	ana	53.0	y33
3	3.0	maria	32.0	x45
4	4.0	fatima	25.0	y78
5	5.0	raul	27.0	w56
6	6.0	julio	18.0	w67
7	7.0	dino		x89
8	8.0	flavia	23.0	x12
9	9.0	jeff	157.0	y45
10	10.0	miriam	56.0	z45

Figura 1: Tabela clientes

No.	1: código Numeric	2: idade Numeric
1	1.0	34.0
2	2.0	53.0
3	3.0	32.0
4	4.0	25.0
5	5.0	27.0
6	6.0	18.0
7	7.0	
8	8.0	23.0
9	9.0	157.0
10	10.0	56.0

Figura 2: Pós retirada de campos.

1. Em relação a tabela de clientes, retiramos os campos de nome e endereço por não serem relevantes para a pesquisa.

No.	1: codigo Numeric	2: idade Numeric
1	1.0	34.0
2	2.0	53.0
3	3.0	32.0
4	4.0	25.0
5	5.0	27.0
6	6.0	18.0
7	7.0	
8	8.0	23.0
9	9.0	157.0
10	10.0	56.0

Figura 3: Tabela Clientes com anomalia

No.	1: codigo Numeric	2: idade Numeric
1	1.0	34.0
2	2.0	53.0
3	3.0	32.0
4	4.0	25.0
5	5.0	27.0
6	6.0	18.0
7	8.0	23.0
8	9.0	157.0
9	10.0	56.0

Figura 4: Pós retirada da anomalia.

2. Retiramos as anomalias da tabela clientes para continuar o pré-processamento.
3. Retiramos os campos Preço, total e hora da tabela de Vendas.
4. Retiramos as tuplas relacionadas ao cliente anomalia 7.

No.	1: codigo Numeric	2: idade Numeric
1	1.0	34.0
2	2.0	53.0
3	3.0	32.0
4	4.0	25.0
5	5.0	27.0
6	6.0	18.0
7	8.0	23.0
8	9.0	157.0
9	10.0	56.0

Figura 5 com Outliers.

No.	1: codigo Numeric	2: idade Numeric
1	1.0	34.0
2	2.0	53.0
3	3.0	32.0
4	4.0	25.0
5	5.0	27.0
6	6.0	18.0
7	8.0	23.0
8	10.0	56.0

Figura 6 sem Outliers.

5. Retiramos os Outliers da tabela clientes para continuar o pré-processamento.

No.	1: codigo Numeric	2: nome Nominal	3: secao Nominal
1	1.0	nabo	legumes
2	2.0	odd	limpeza
3	3.0	carne 1a	carnes
4	4.0	cenoura	legumes
5	5.0	dorito	salgadi...
6	6.0	algodao	gerais
7	7.0	alface	verduras
8	8.0	linguica	carnes
9	9.0	tomate	legumes
10	10.0	bombril	limpeza

Figura 7: Tabela Produtos.

No.	1: codigo Numeric	2: nome Nominal	3: secao Nominal
1	1.0	nabo	legumes
2	3.0	carne 1a	carnes
3	4.0	cenoura	legumes
4	5.0	dorito	salgadinhos
5	7.0	alface	verduras
6	8.0	linguica	carnes
7	9.0	tomate	legu... ▼

Figura 8: Tabela Produtos sem seção gerais e limpeza.

6. Retiramos os produtos das seções de limpeza e gerais (Algodão, Odd e Bombril).

No.	1: codigovenda Numeric	2: data Nominal	3: hora Nominal	4: cliente Numeric	5: produto Numeric	6: qtde Numeric	7: preco Numeric	8: total Numeric
1	1.0	01/01...	08:01	4.0	8.0	1.0	1.54	1.54
2	2.0	02/01...	08:02	4.0	9.0	1.0	1.96	1.96
3	3.0	03/01...	08:03	4.0	1.0	2.0	1.57	3.14
4	4.0	04/01...	08:04	6.0	3.0	2.0	1.29	2.58
5	5.0	05/01...	08:05	0.0	5.0	3.0	1.64	4.92
6	6.0	06/01...	08:06	6.0	9.0	3.0	1.73	5.19
7	7.0	07/01...	08:07	6.0	4.0	1.0	1.6	1.6
8	8.0	08/01...	08:08	8.0	9.0	2.0	1.47	2.94
9	9.0	09/01...	08:09	1.0	2.0	3.0	1.47	4.41
10	10.0	10/01...	08:10	4.0	5.0	3.0	1.65	4.95
11	11.0	11/01...	08:11	8.0	8.0	2.0	1.92	3.84
12	12.0	12/01...	08:12	7.0	4.0	1.0	1.15	1.15
13	13.0	13/01...	08:13	1.0	8.0	3.0	1.65	4.95
14	14.0	14/01...	08:14	3.0	8.0	1.0	1.22	1.22
15	15.0	15/01...	08:15	10.0	5.0	3.0	1.88	5.64
16	16.0	16/01...	08:16	3.0	4.0	2.0	1.85	3.7
17	17.0	17/01...	08:17	1.0	3.0	2.0	1.34	2.68
18	18.0	18/01...	08:18	0.0	6.0	3.0	1.38	4.14
19	19.0	19/01...	08:19	4.0	1.0	2.0	1.78	3.56
20	20.0	20/01...	08:20	3.0	5.0	1.0	1.37	1.37
21	21.0	21/01...	08:21	7.0	6.0	3.0	1.43	4.29
22	22.0	22/01...	08:22	0.0	9.0	2.0	1.52	3.04
23	23.0	23/01...	08:23	7.0	9.0	2.0	1.38	2.76
24	24.0	24/01...	08:24	4.0	5.0	2.0	1.19	2.38
25	25.0	25/01...	08:25	10.0	9.0	1.0	1.33	1.33
26	26.0	26/01...	08:26	3.0	6.0	1.0	1.01	1.01
27	27.0	27/01...	08:27	9.0	5.0	3.0	1.76	5.28
28	28.0	28/01...	08:28	5.0	4.0	1.0	1.55	1.55
29	29.0	29/01...	08:29	5.0	7.0	1.0	1.21	1.21
30	30.0	30/01...	08:30	6.0	6.0	2.0	1.11	2.22
31	31.0	31/01...	08:31	5.0	1.0	1.0	1.99	1.99

Figura 9: Tabela de Vendas.

No.	1: cliente Numeric	2: produto Numeric
1	4.0	8.0
2	4.0	9.0
3	4.0	1.0
4	6.0	3.0
5	0.0	5.0
6	6.0	9.0
7	6.0	4.0
8	8.0	9.0
9	1.0	2.0
10	4.0	5.0
11	8.0	8.0
12	1.0	8.0
13	3.0	8.0
14	10.0	5.0
15	3.0	4.0
16	1.0	3.0
17	0.0	6.0
18	4.0	1.0
19	3.0	5.0
20	0.0	9.0
21	4.0	5.0
22	10.0	9.0
23	3.0	6.0
24	9.0	5.0
25	5.0	4.0
26	5.0	7.0
27	6.0	6.0
28	5.0	1.0

Figura 10: Tabela vendas.

7. Removemos da tabela venda os atributos (Data, Código Venda, Quantidade).

No.	1: cliente Numeric	2: produto Numeric
1	4.0	8.0
2	4.0	9.0
3	4.0	1.0
4	6.0	3.0
5	6.0	9.0
6	6.0	4.0
7	8.0	9.0
8	1.0	2.0
9	4.0	5.0
10	8.0	8.0
11	1.0	8.0
12	3.0	8.0
13	10.0	5.0
14	3.0	4.0
15	1.0	3.0
16	4.0	1.0
17	3.0	5.0
18	4.0	5.0
19	10.0	9.0
20	3.0	6.0
21	5.0	4.0
22	5.0	7.0
23	6.0	6.0
24	5.0	1.0

Figura 10: Tabela vendas sem clientes (0, 7, 9).

8. Removemos os clientes de vendas (0, 7, 9).

9. 0 Por não existir, 7 por ser anomalia na tabela cliente, 9 por ser um outilier na tabela clientes.

No.	1: cliente Numeric	2: produto Numeric
1	4.0	8.0
2	4.0	9.0
3	4.0	1.0
4	8.0	9.0
5	1.0	2.0
6	4.0	5.0
7	8.0	8.0
8	1.0	8.0
9	3.0	8.0
10	3.0	4.0
11	1.0	3.0
12	4.0	1.0
13	3.0	5.0
14	4.0	5.0
15	3.0	6.0
16	5.0	4.0
17	5.0	7.0
18	5.0	1.0

Figura 11: Tabela vendas sem produtos (2, 6, 10).

10. Removemos os códigos de produtos 2, 6 e 10.

2 - Algodão/Gerais, queremos apenas produtos Alimentício.

6 - Odd/Limpeza, queremos apenas produtos Alimentício.

10 – Bombril/Limpeza, queremos apenas produtos Alimentício.

No.	1: cliente Numeric	2: produto Numeric	3: idade Numeric	4: nomeProduto Nominal	5: secao Nominal
1	4.0	8.0	25.0	linguica	carnes
2	4.0	9.0	25.0	tomate	legumes
3	4.0	1.0	25.0	tomate	legumes
4	8.0	9.0	23.0	linguica	carnes
5	1.0	2.0	34.0	nabo	legumes
6	4.0	5.0	25.0	cenoura	legumes
7	8.0	8.0	23.0	linguica	carnes
8	1.0	8.0	34.0	nabo	legumes
9	3.0	8.0	32.0	carne 1a	carnes
10	3.0	4.0	32.0	carne 1a	carnes
11	1.0	3.0	34.0	nabo	legumes
12	4.0	1.0	25.0	cenoura	legumes
13	3.0	5.0	32.0	carne 1a	carnes
14	4.0	5.0	25.0	cenoura	legumes
15	3.0	6.0	32.0	carne 1a	carnes
16	5.0	4.0	27.0	dorito	salgadinhos
17	5.0	7.0	27.0	dorito	salgadinhos
18	5.0	1.0	27.0	dorito	salgadinhos

Figura 12: Tabela vendas com atributos novos.

11. Criamos novos atributos (Idade, NomeProduto, Seção)

No.	1: idade Numeric	2: nomeProduto Nominal	3: secao Nominal
1	25.0	linguica	carnes
2	25.0	tomate	legumes
3	25.0	tomate	legumes
4	23.0	linguica	carnes
5	34.0	nabo	legumes
6	25.0	cenoura	legumes
7	23.0	linguica	carnes
8	34.0	nabo	legumes
9	32.0	carne 1a	carnes
10	32.0	carne 1a	carnes
11	34.0	nabo	legumes
12	25.0	cenoura	legumes
13	32.0	carne 1a	carnes
14	25.0	cenoura	legumes
15	32.0	carne 1a	carnes
16	27.0	dorito	salgadinhos
17	27.0	dorito	salgadinhos
18	27.0	dorito	salgadinhos

Figura 12: Tabela Vendas sem Id Cliente e Produto.

12. Removemos os atributos (Cliente e Produtos)

No.	1: idade Nominal	2: Nome Nominal	3: secao Nominal
1	'(-inf-26.666667]'	linguica	carnes
2	'(-inf-26.666667]'	linguica	carnes
3	'(-inf-26.666667]'	linguica	legumes
4	'(-inf-26.666667]'	tomate	legumes
5	'(-inf-26.666667]'	tomate	legumes
6	'(-inf-26.666667]'	cenoura	legumes
7	'(-inf-26.666667]'	cenoura	legumes
8	'(-inf-26.666667]'	cenoura	legumes
9	'(26.666667-30.333333]'	dorito	salgadi...
10	'(26.666667-30.333333]'	dorito	salgadi...
11	'(26.666667-30.333333]'	dorito	salgadi...
12	'(30.333333-inf)'	nabo	legumes
13	'(30.333333-inf)'	nabo	legumes
14	'(30.333333-inf)'	nabo	legumes
15	'(30.333333-inf)'	'carne 1a'	carnes
16	'(30.333333-inf)'	'carne 1a'	carnes
17	'(30.333333-inf)'	'carne 1a'	carnes
18	'(30.333333-inf)'	'carne 1a'	carnes

Figura 12: Tabela Vendas com discretize na idade.

13. Usamos discretize na idade.
 - 1 – Faixa (-inf , 26.66)
Coseguimos obter 8 registros.
 - 2 – Faixa (26.66, -30.33)
Conseguimos obter 3 registros.
 - 3 – Faixa (30.33, inf)
Conseguimos obter 7 registros.

Base de dados

Arquivo .csv

```
idade,nomeProduto,secao
"\'(-inf-26.666667]\",linguica,legumes
"\'(-inf-26.666667]\",tomate,legumes
"\'(-inf-26.666667]\",tomate,legumes
"\'(-inf-26.666667]\",linguica,carnes
"\'(30.333333-inf)\",nabo,legumes
"\'(-inf-26.666667]\",cenoura,legumes
"\'(-inf-26.666667]\",linguica,carnes
"\'(30.333333-inf)\",nabo,legumes
"\'(30.333333-inf)\",\'carne 1a',carnes
"\'(30.333333-inf)\",\'carne 1a',carnes
"\'(30.333333-inf)\",nabo,legumes
"\'(-inf-26.666667]\",cenoura,legumes
"\'(30.333333-inf)\",\'carne 1a',carnes
"\'(-inf-26.666667]\",cenoura,legumes
"\'(30.333333-inf)\",\'carne 1a',carnes
"\'(26.666667-30.333333]\",dorito,salgadinhos
"\'(26.666667-30.333333]\",dorito,salgadinhos
"\'(26.666667-30.333333]\",dorito,salgadinhos
```