



Universidade Estadual do Oeste do Paraná  
Disciplina de Inteligência Artificial  
Docentes: Huei D. Lee e Newton Spolaor

# Mineração de dados

## Projeto 4

Discentes:

Isabela Loebel

Nickolas Crema



# Domínio da aplicação



## *Estimation of Obesity Levels Based On Eating Habits and Physical Condition [1];*

(Estimativa dos níveis de obesidade com base nos hábitos alimentares e na condição física – Tradução livre)

- *Fabio Mendoza Palechor e Alexis De la Hoz Manotas, 2019;*
- *Países do estudo: México, Peru e Colômbia;*
- *16 atributos, 1 classe:*
  - *8 atributos numéricos;*
  - *8 atributos categóricos.*
- *2.116 casos;*

# Atributos

## Númericos:

- Idade;
- Peso;
- Altura;
- Come vegetais nas refeições;
- Quantidade de refeições diárias;
- Consumo de água diário;
- Atividades físicas semanais;
- Tempo gasto com dispositivos tecnológicos.

## Categóricos:

- Gênero;
- Fuma;
- Come entre as refeições;
- Histórico de sobrepeso familiar;
- Monitora calorias;
- Come comida muito calórica;
- Frequência de consumo de álcool;
- Meio de transporte.



# Objetivo

- Estimar o nível de obesidade de um indivíduo com base na sua condição física e hábitos saudáveis;
- Classificações:
  - *Abaixo do Peso;*
  - *Peso Normal;*
  - *Sobrepeso Nível I;*
  - *Sobrepeso Nível II;*
  - *Obesidade Tipo I;*
  - *Obesidade Tipo II;*
  - *Obesidade Tipo III;*



# Primeiro contato com os dados

- Classe bem distribuída;
- Dataset balanceado por meio de SMOTE pelo criador.

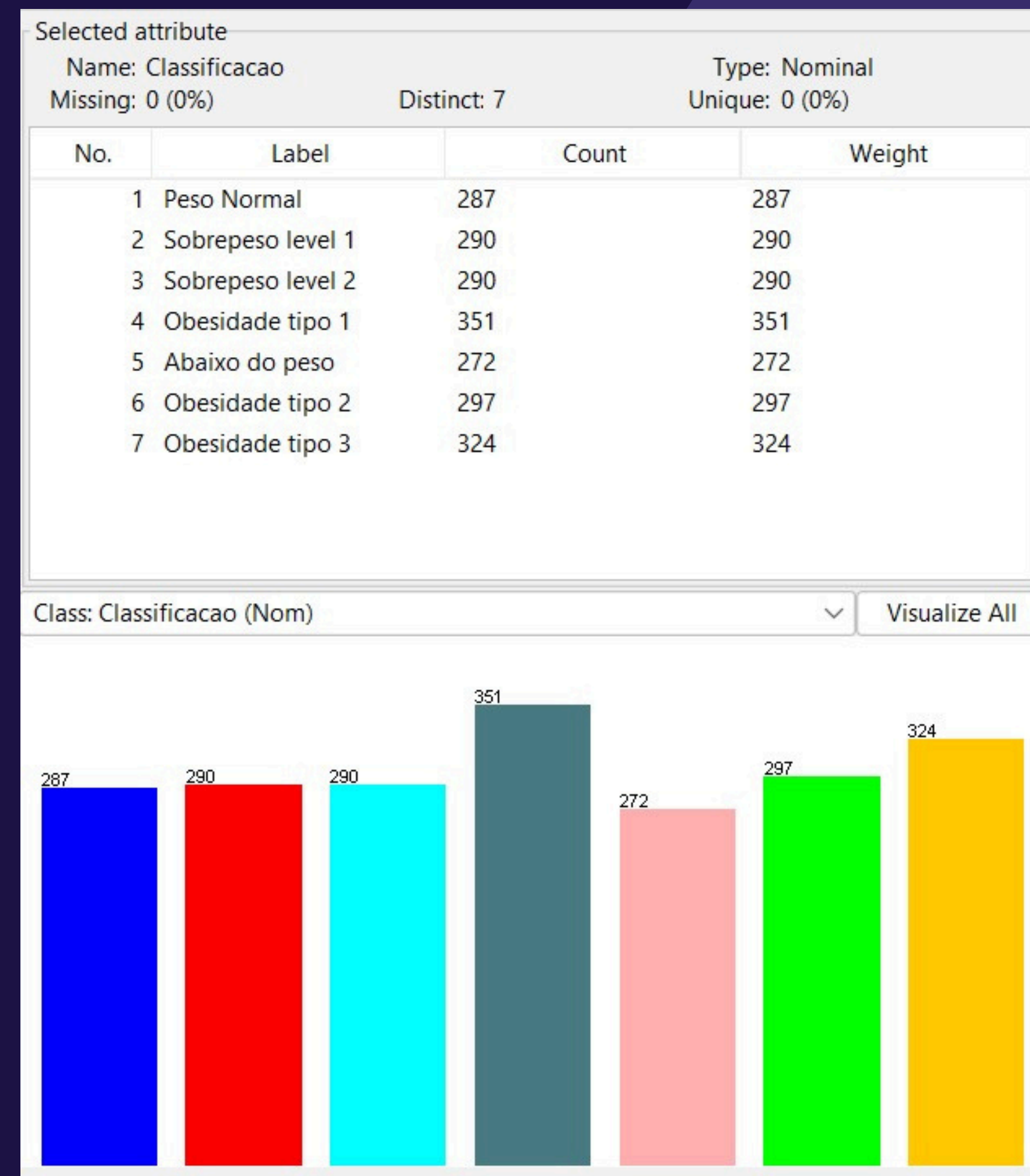


Figura 1 – Distribuição da classe. Fonte: Autores.



# Pré-processamento

Atributo: Come vegetais nas refeições.

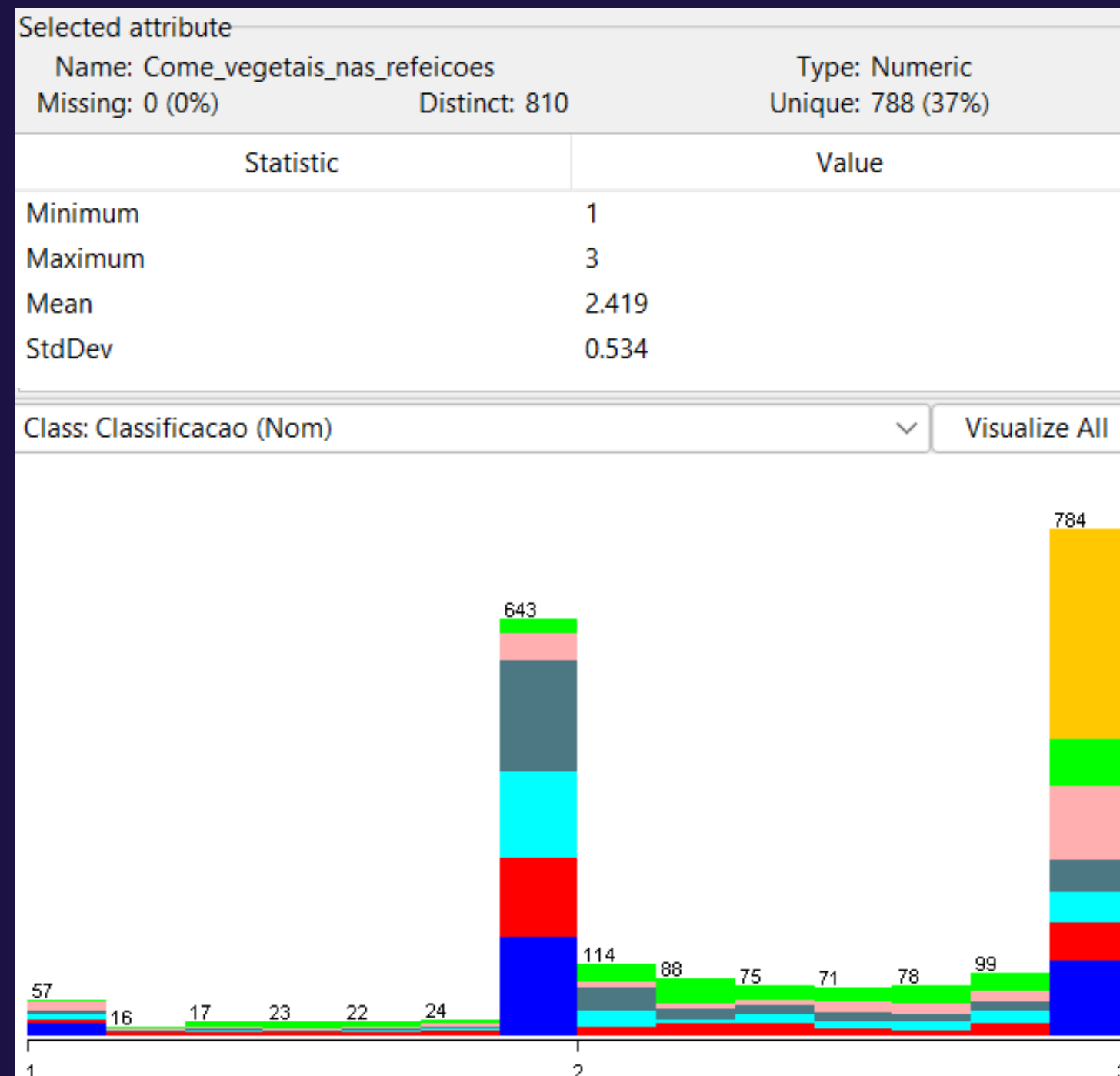


Figura 3 - Atributo "come vegetais nas refeições" discretizado. Fonte: Autores.

Figura 2 - Atributo "come vegetais nas refeições" numérico. Fonte: Autores.

# Pré-processamento

Atributo: Consumo de água diário.



Figura 4 - Atributo "consumo de água diário" numérico. Fonte: Autores.

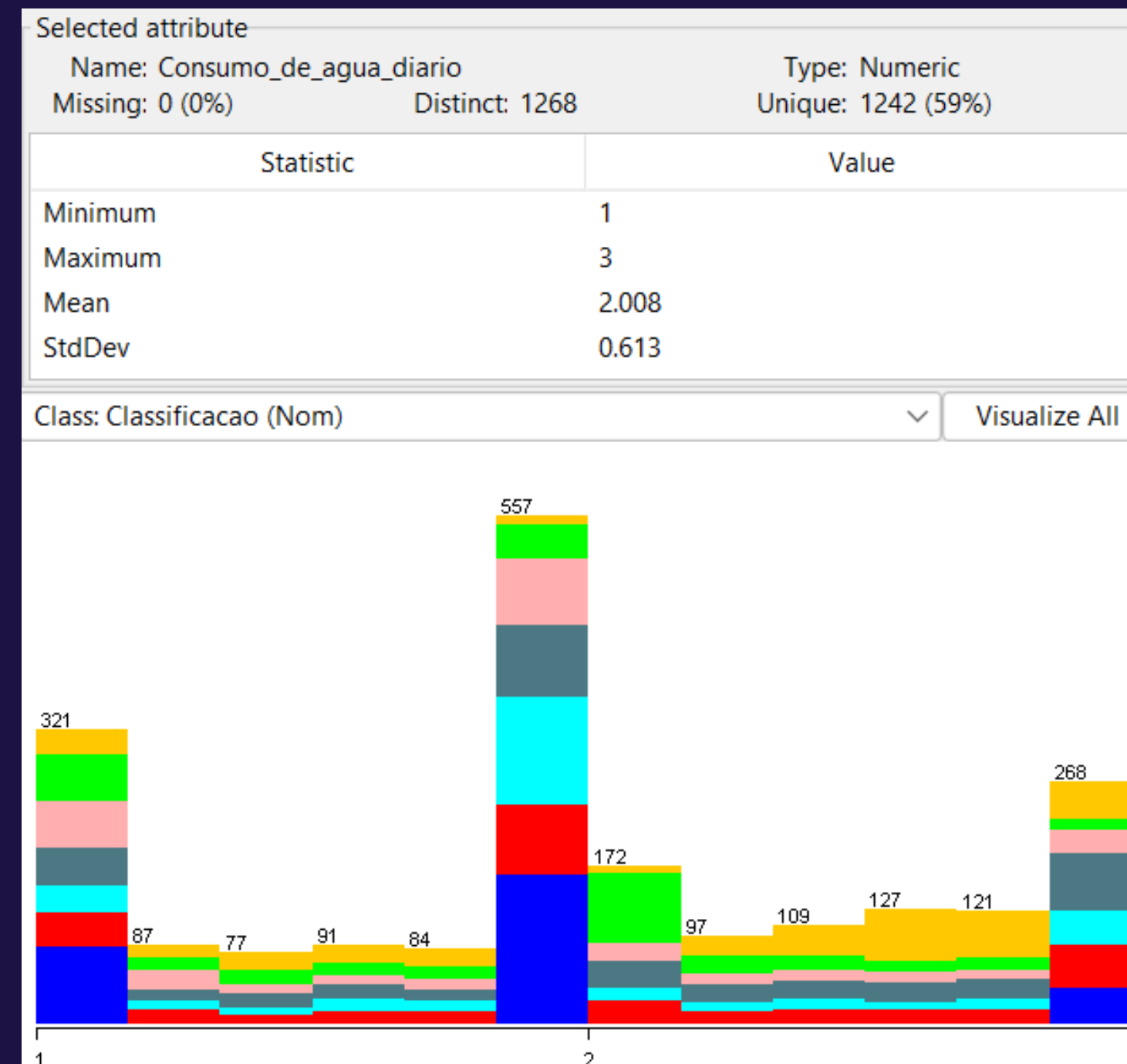


Figura 5 - Atributo "consumo de água diário" discretizado. Fonte: Autores.

# Pré-processamento

Atributo: Frequência de atividade física.

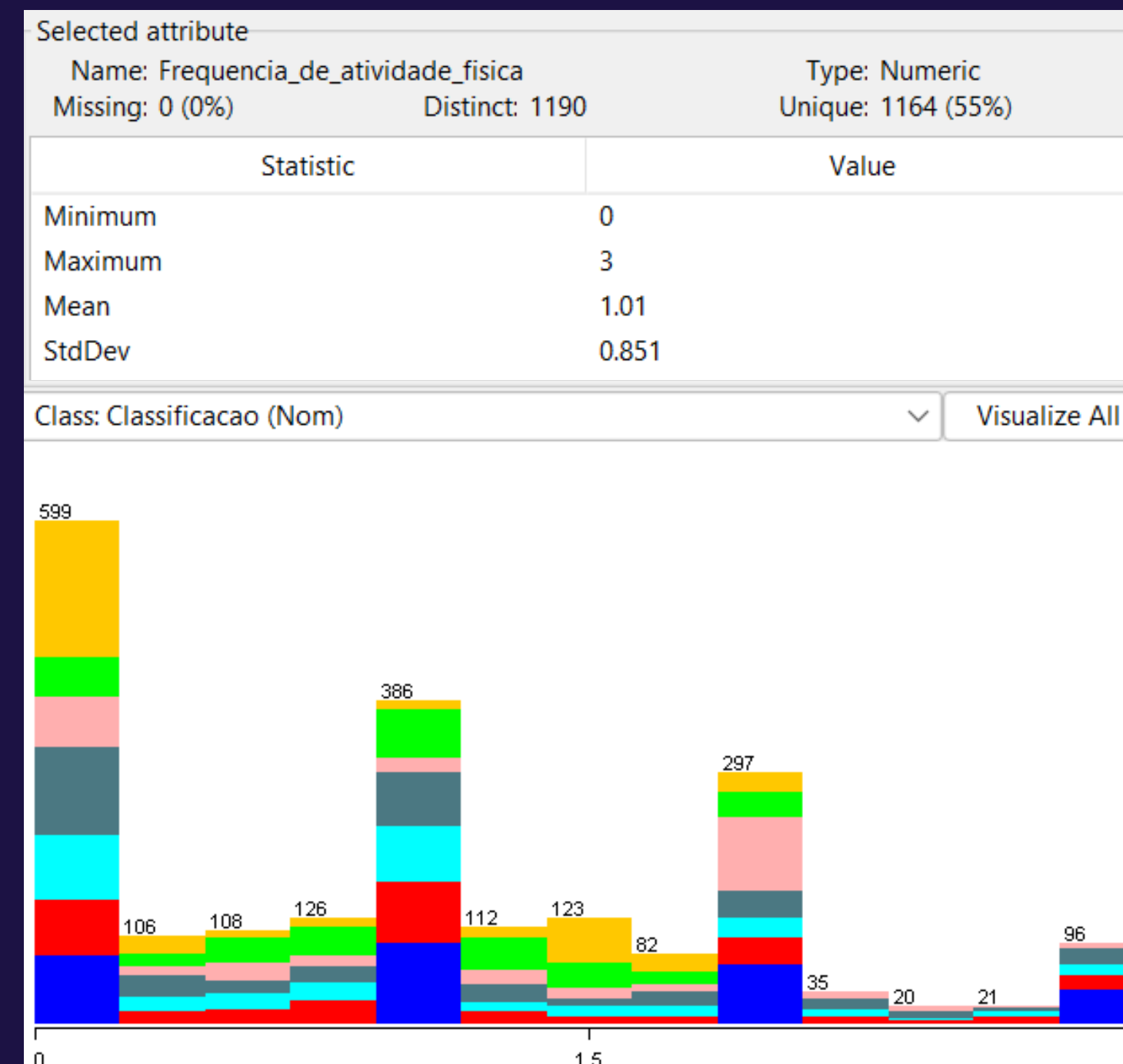
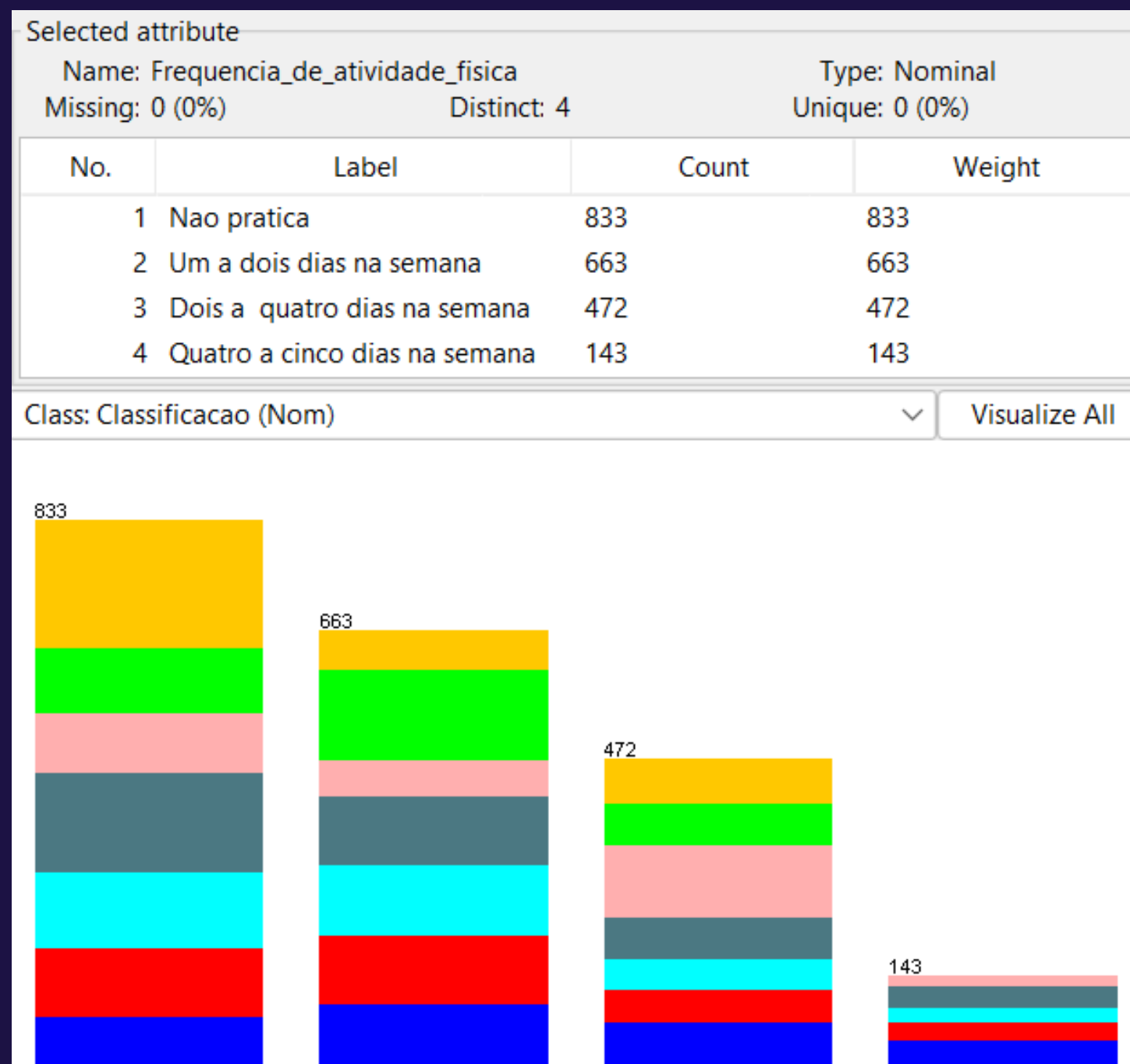


Figura 6 – Atributo “consumo de água diário” numérico. Fonte: Autores.

Figura 7 – Atributo “consumo de água diário” discretizado. Fonte: Autores.



# Pré-processamento

Atributo: Quantidade de refeições diárias.

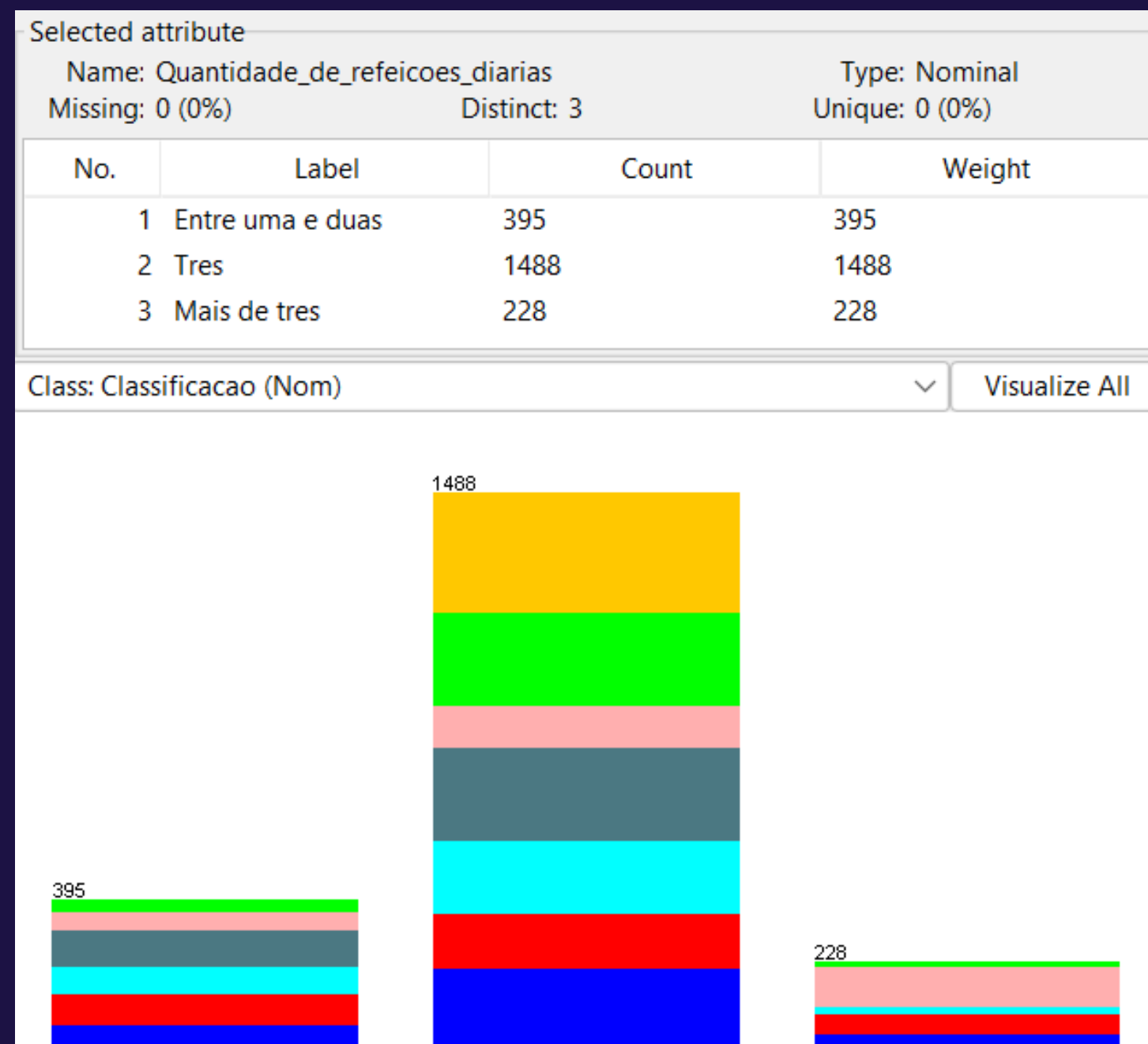


Figura 8 - Atributo "quantidade de refeições diárias" numérico.  
Fonte: Autores.

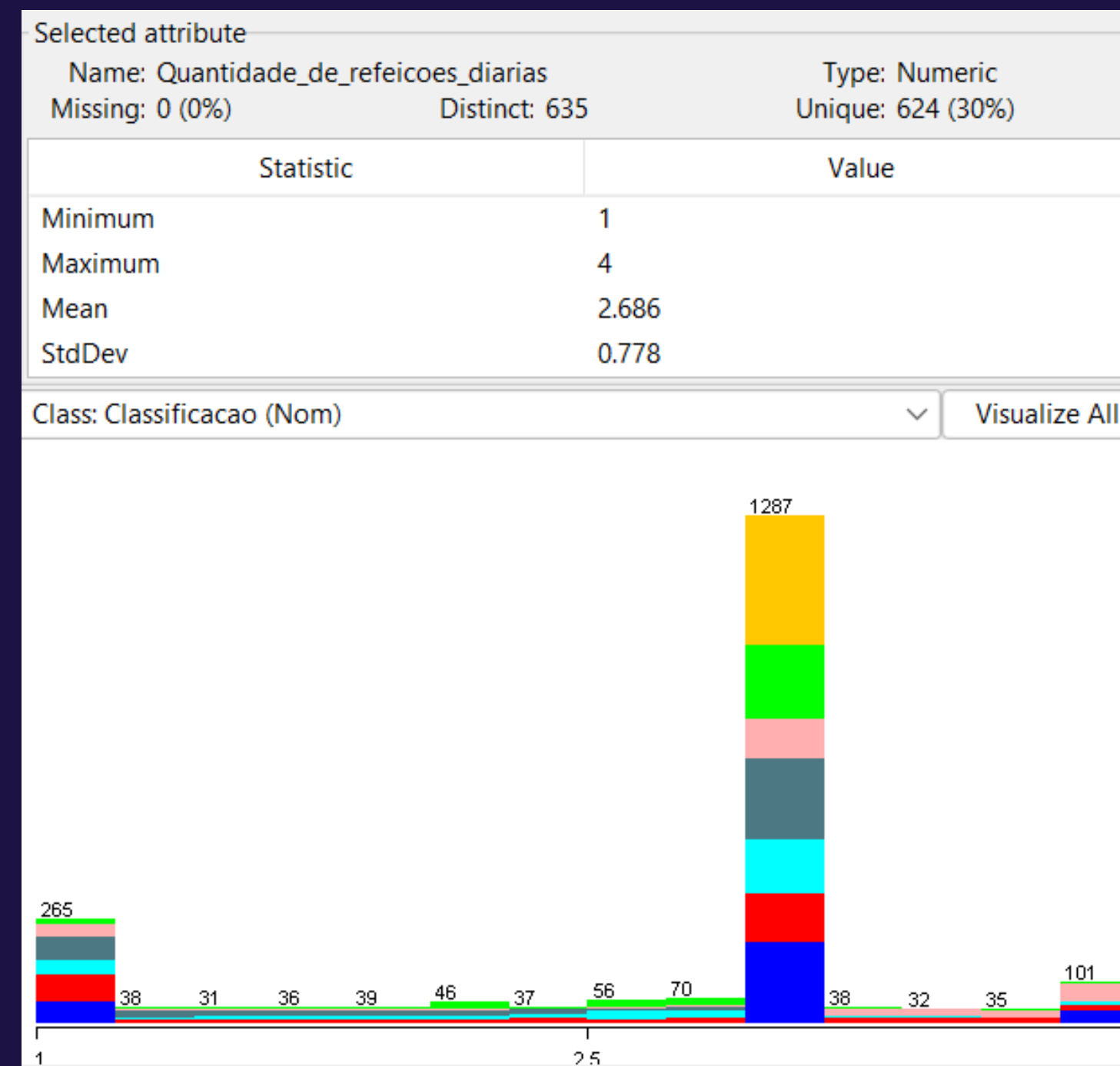


Figura 9 - Atributo "quantidade de refeições diárias" discretizado.  
Fonte: Autores.

# Pré-processamento

## Atributo: IMC.

O **IMC**, ou Índice de Massa Corporal, indica se a pessoa está com um peso **ideal** ou se está **abaixo** ou **acima** do peso. Esta é uma medida de referência internacional reconhecida pela **OMS** (Organização Mundial da Saúde) [3].

$$\text{IMC} = \frac{\text{Peso}}{\text{Altura}^2}$$

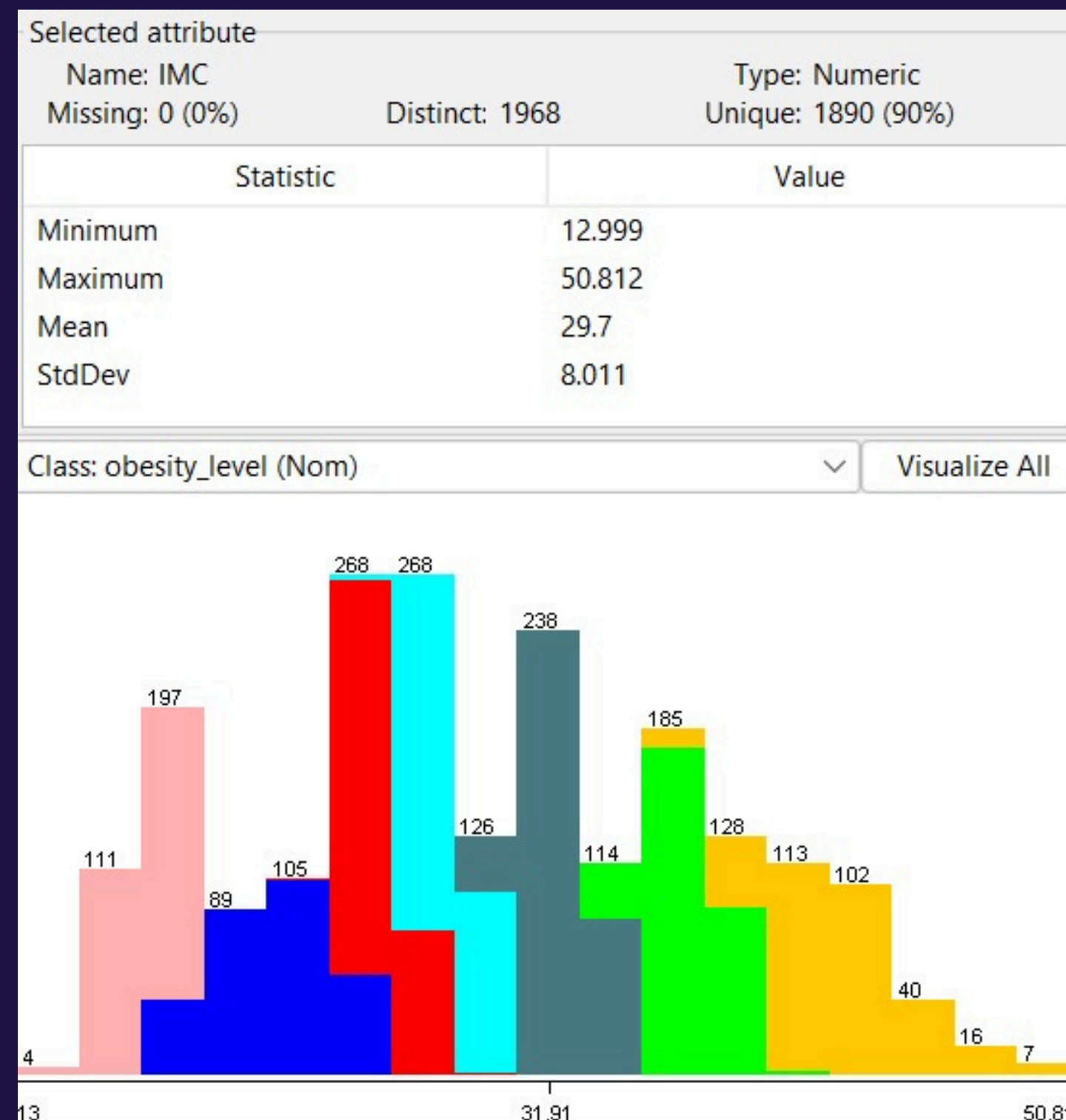


Figura 10 – Atributo “IMC”. Fonte: Autores.

# Pré-processamento

## Seleção de Atributos

InfoGainAttributeEval  
(*Ganho de Informação*)



```
=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===
```

average merit	average rank	attribute
2.624 +- 0.004	1 +- 0	3 IMC
0.801 +- 0.01	2 +- 0	2 Idade
0.298 +- 0.004	3 +- 0	1 Genero_de_Nascimento
0.233 +- 0.005	4.6 +- 0.8	8 Come_entre_as_refeicoes
0.23 +- 0.006	4.9 +- 0.7	4 Tem_historico_de_sobrepeso_familiar
0.227 +- 0.003	5.5 +- 0.67	6 Come_vegetais_nas_refeicoes
0.201 +- 0.004	7 +- 0	7 Quantidade_de_refeicoes_diarias
0.147 +- 0.004	8 +- 0	14 Frequencia_de_consumo_de_alcool
0.112 +- 0.002	9 +- 0	15 Meio_de_transporte
0.088 +- 0.003	10.3 +- 0.46	12 Frequencia_de_atividade_fisica
0.087 +- 0.003	10.7 +- 0.46	5 Come_comidas_caloricas_com_frequencia
0.076 +- 0.004	12.2 +- 0.4	10 Consumo_de_agua_diario
0.072 +- 0.002	12.8 +- 0.4	13 Tempo_gasto_com_dispositivos_tecnologicos
0.046 +- 0.001	14 +- 0	11 Monitora_as_calorias_ingeridas
0.011 +- 0.002	15 +- 0	9 Fuma

Figura 11 - Resultado da seleção de atributos. Fonte: Autores.

# Mineração de Dados

Cenários:

Algoritmo J48Tree

1

Dados pré-processados.

2

Dados pré-processados com inclusão do atributo IMC.

3

Dados pré-processados com inclusão do atributo IMC e seleção de atributos.

# Mineração de Dados

## Cenário 1

Número de Folhas

65

Tamanho da Árvore

121

## Acurácia da classe

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,880	0,033	0,826	0,880	0,852	0,825	0,951	0,778	Peso Normal
	0,870	0,024	0,842	0,870	0,856	0,834	0,948	0,808	Sobrepeso level 1
	0,887	0,014	0,905	0,887	0,896	0,880	0,958	0,867	Sobrepeso level 2
	0,927	0,010	0,950	0,927	0,938	0,926	0,988	0,929	Obesidade tipo 1
	0,910	0,005	0,964	0,910	0,936	0,928	0,996	0,965	Abaixo do peso
	0,971	0,005	0,971	0,971	0,971	0,966	0,988	0,971	Obesidade tipo 2
	0,990	0,002	0,990	0,990	0,990	0,989	0,994	0,982	Obesidade tipo 3
Weighted Avg.	0,921	0,013	0,922	0,921	0,921	0,908	0,975	0,901	

Figura 12.

## Matriz de Confusão

a	b	c	d	e	f	g	<-- classified as
95	10	0	0	3	0	0	a = Peso Normal
10	80	2	0	0	0	0	b = Sobrepeso level 1
2	5	86	4	0	0	0	c = Sobrepeso level 2
0	0	7	114	0	2	0	d = Obesidade tipo 1
8	0	0	0	81	0	0	e = Abaixo do peso
0	0	0	2	0	101	1	f = Obesidade tipo 2
0	0	0	0	0	1	104	g = Obesidade tipo 3

Figura 13.

## Sumário

Correctly Classified Instances	661	92.0613 %
Incorrectly Classified Instances	57	7.9387 %
Kappa statistic	0.9072	
Mean absolute error	0.0282	
Root mean squared error	0.1457	
Relative absolute error	11.5127 %	
Root relative squared error	41.6385 %	
Total Number of Instances	718	

Figura 14.



# Mineração de Dados

## Cenário 2

Número de Folhas

29

Tamanho da Árvore

52

## Acurácia da classe

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	0,007	0,964	1,000	0,982	0,979	0,996	0,961	Peso Normal
	0,946	0,003	0,978	0,946	0,961	0,956	0,992	0,948	Sobrepeso level 1
	0,979	0,010	0,941	0,979	0,960	0,953	0,993	0,927	Sobrepeso level 2
	0,967	0,003	0,983	0,967	0,975	0,970	0,981	0,973	Obesidade tipo 1
	0,966	0,000	1,000	0,966	0,983	0,981	0,997	0,982	Abaixo do peso
	0,962	0,003	0,980	0,962	0,971	0,966	0,988	0,957	Obesidade tipo 2
	0,990	0,005	0,972	0,990	0,981	0,978	0,993	0,964	Obesidade tipo 3
Weighted Avg.	0,974	0,004	0,974	0,974	0,974	0,969	0,991	0,959	

Figura 15.

## Matriz de Confusão

	a	b	c	d	e	f	g	<-- classified as
108	0	0	0	0	0	0	0	a = Peso Normal
1	87	4	0	0	0	0	0	b = Sobrepeso level 1
0	2	95	0	0	0	0	0	c = Sobrepeso level 2
0	0	2	119	0	1	1	1	d = Obesidade tipo 1
3	0	0	0	0	86	0	0	e = Abaixo do peso
0	0	0	2	0	100	2	2	f = Obesidade tipo 2
0	0	0	0	0	0	1	104	g = Obesidade tipo 3

Figura 16.

## Sumário

Correctly Classified Instances	699	97.3538 %
Incorrectly Classified Instances	19	2.6462 %
Kappa statistic	0.9691	
Mean absolute error	0.0113	
Root mean squared error	0.0855	
Relative absolute error	4.6342 %	
Root relative squared error	24.4493 %	
Total Number of Instances	718	

Figura 17.

# Mineração de Dados

## Cenário 3

Número de Folhas

15

Tamanho da Árvore

28

## Acurácia da classe

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	0,010	0,947	1,000	0,973	0,969	0,995	0,947	Peso Normal
	0,924	0,003	0,977	0,924	0,950	0,943	0,987	0,933	Sobrepeso level 1
	0,969	0,008	0,949	0,969	0,959	0,953	0,992	0,934	Sobrepeso level 2
	0,984	0,003	0,984	0,984	0,984	0,980	0,990	0,978	Obesidade tipo 1
	0,966	0,000	1,000	0,966	0,983	0,981	0,997	0,981	Abaixo do peso
	0,971	0,002	0,990	0,971	0,981	0,977	0,989	0,974	Obesidade tipo 2
	0,990	0,005	0,972	0,990	0,981	0,978	0,993	0,964	Obesidade tipo 3
Weighted Avg.	0,974	0,005	0,974	0,974	0,973	0,969	0,992	0,959	

Figura 18.

## Matriz de Confusão

a	b	c	d	e	f	g	<-- classified as
108	0	0	0	0	0	0	a = Peso Normal
3	85	4	0	0	0	0	b = Sobrepeso level 1
0	2	94	1	0	0	0	c = Sobrepeso level 2
0	0	1	121	0	0	1	d = Obesidade tipo 1
3	0	0	0	86	0	0	e = Abaixo do peso
0	0	0	1	0	101	2	f = Obesidade tipo 2
0	0	0	0	0	1	104	g = Obesidade tipo 3

Figura 19.

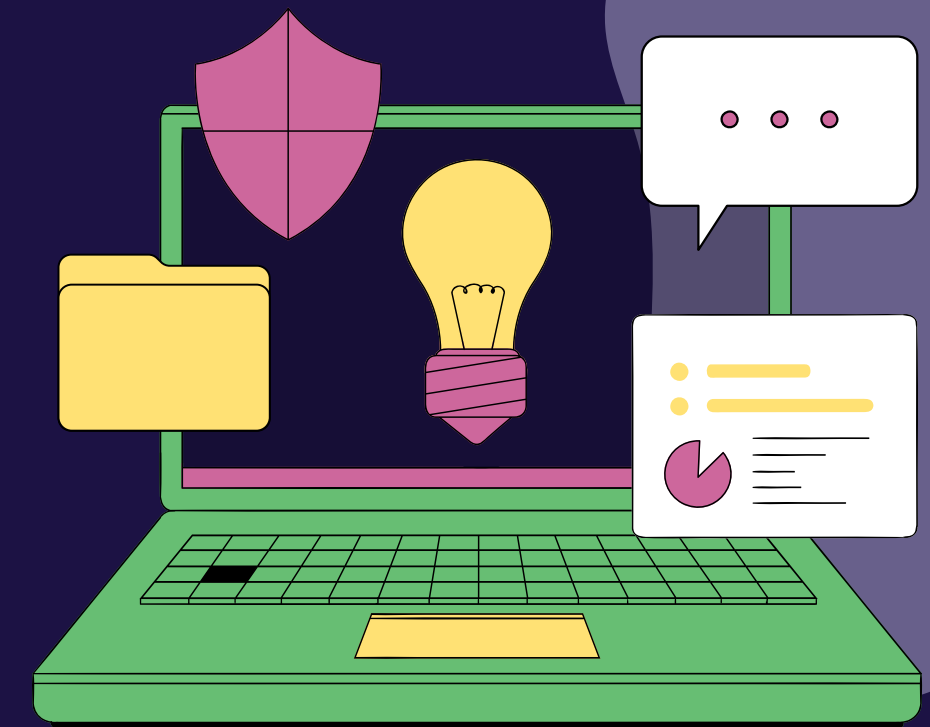
## Sumário

Correctly Classified Instances	699	97.3538 %
Incorrectly Classified Instances	19	2.6462 %
Kappa statistic	0.9691	
Mean absolute error	0.0127	
Root mean squared error	0.0859	
Relative absolute error	5.1718 %	
Root relative squared error	24.5521 %	
Total Number of Instances	718	

Figura 20.

# Dificuldades Encontradas

- Devido ao *dataset* já ter sido aplicado SMOTE, restringiu o estudo às escolhas que os autores estabeleceram;
- Dificuldade no tratamento de dados que eram categóricos e passaram a ser numéricos.



# Referências Bibliográficas

- [1] PALECHOR, F. M.; MANOTAS, A. De la H. **Estimation of Obesity Levels Based On Eating Habits and Physical Condition** . (2019). UCI Machine Learning Repository. Disponível em: <<https://doi.org/10.24432/C5H31Z>>. Acesso em março de 2024.
- [2] PALECHOR, F. M.; MANOTAS, A. De la H. **“Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico.”** 2019. National Library of Medicine – National Center for Biotechnology Information.
- [3] ANJOS, L. A. **Índice de massa corporal (massa corporal.estatura-2) como indicador do estado nutricional de adultos: revisão da literatura.** Centro de Estudo da Saúde do Trabalhador e Ecologia Humana – Escola Nacional de Saúde Pública – Fundação Oswaldo Cruz. Rev. Saúde Pública vol. 26 n. 6. 1992. Rio de Janeiro, RJ – Brasil.



# Obrigado!

## Dúvidas?



Contatos:

Isabela — [isabelaloebel@gmail.com](mailto:isabelaloebel@gmail.com)

Nickolas — [nick.cremaa@gmail.com](mailto:nick.cremaa@gmail.com)