



# Projeto de Machine Learning - Etapa 2

Análise de sentimento de reviews

Larissa Lewartoski Wong

# O Conjunto de Dados



## Amazon Fine Food Reviews

Analyze ~500,000 food reviews from Amazon



### Atributos

- ProductId - Identificador do produto;
- UserId - Identificador do usuário;
- ProfileName - Nickname do usuário;
- HelpfulnessNumerator - Número de marcações como útil;
- HelpfulnessDenominator - Número de marcações geral, útil e inútil;
- Score - Avaliação de 1 a 5;
- Time - Timestamp do review;
- Summary - Sumário curto do review;
- Text - Texto completo do review.

# Análise exploratória dos dados

Análise dos tipos e verificação de nulos

Id	0
ProductId	0
UserId	0
ProfileName	16
HelpfulnessNumerator	0
HelpfulnessDenominator	0
Score	0
Time	0
Summary	27
Text	0

RangeIndex: 568454 entries, 0 to 568453

Data columns (total 10 columns):

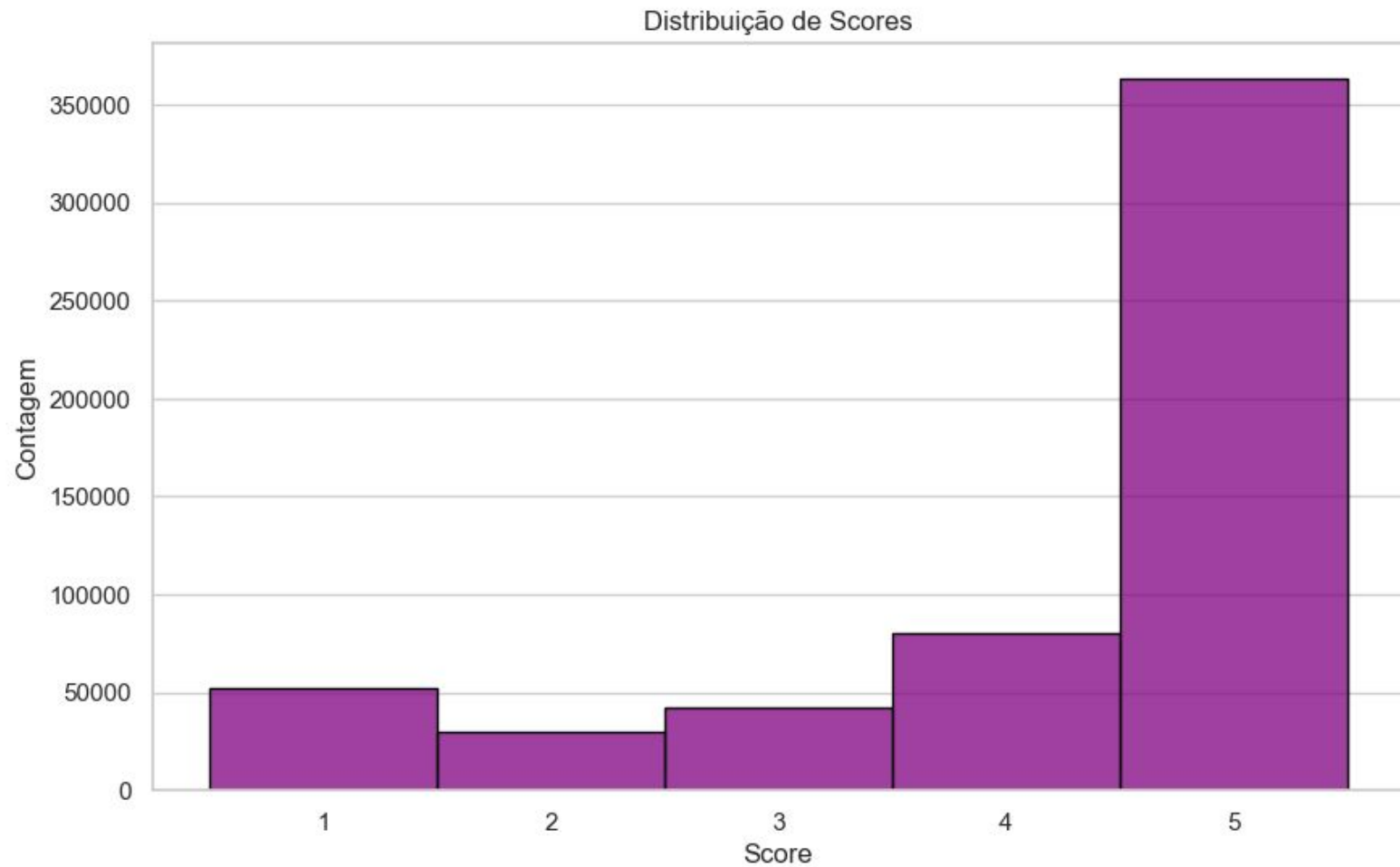
#	Column	Non-Null Count	Dtype
0	Id	568454 non-null	int64
1	ProductId	568454 non-null	object
2	UserId	568454 non-null	object
3	ProfileName	568438 non-null	object
4	HelpfulnessNumerator	568454 non-null	int64
5	HelpfulnessDenominator	568454 non-null	int64
6	Score	568454 non-null	int64
7	Time	568454 non-null	int64
8	Summary	568427 non-null	object
9	Text	568454 non-null	object

dtypes: int64(5), object(5)

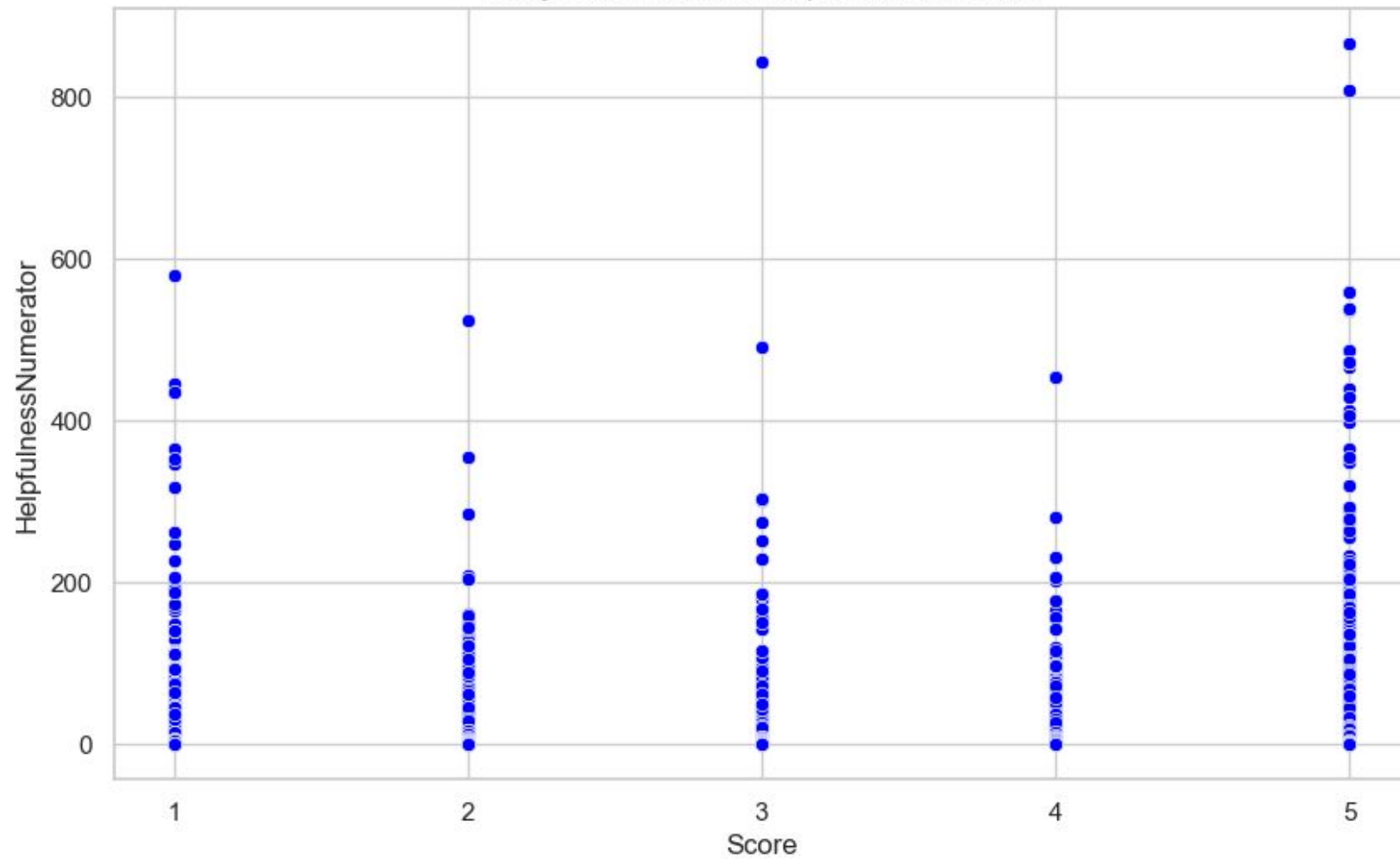
# Análise das variáveis numéricas

- A maior parte dos reviews possui score entre 4 e 5 ;
- Marcações de usuários quanto à utilidade dos reviews é baixa.

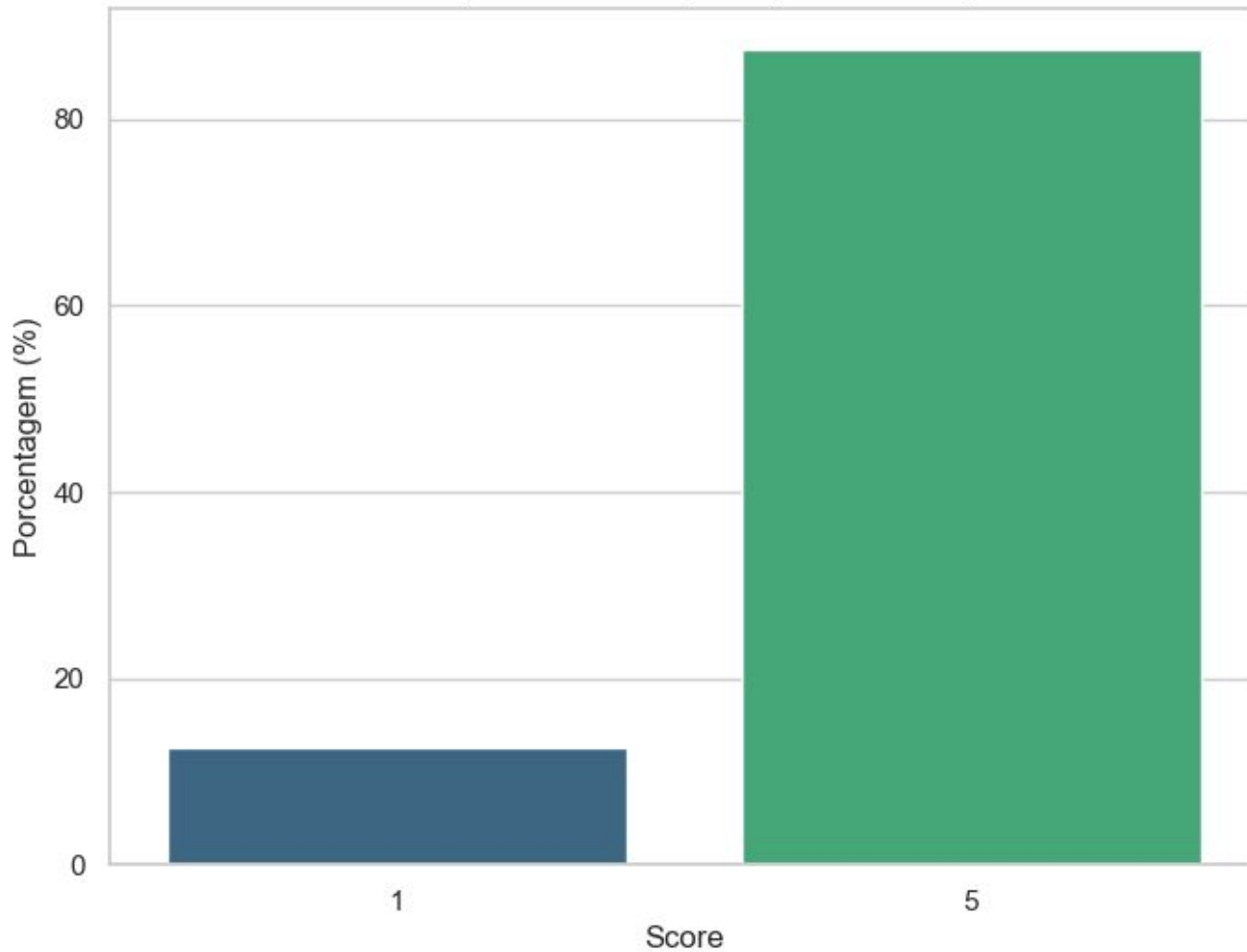
	<b>Id</b>	<b>HelpfulnessNumerator</b>	<b>HelpfulnessDenominator</b>	<b>Score</b>	<b>Time</b>
count	568454.000000	568454.000000	568454.00000	568454.000000	5.684540e+05
mean	284227.500000	1.743817	2.22881	4.183199	1.296257e+09
std	164098.679298	7.636513	8.28974	1.310436	4.804331e+07
min	1.000000	0.000000	0.00000	1.000000	9.393408e+08
25%	142114.250000	0.000000	0.00000	4.000000	1.271290e+09
50%	284227.500000	0.000000	1.00000	5.000000	1.311120e+09
75%	426340.750000	2.000000	2.00000	5.000000	1.332720e+09
max	568454.000000	866.000000	923.00000	5.000000	1.351210e+09



Relação entre Score e HelpfulnessNumerator



Distribuição de Classes (Score) em Porcentagem



Score 5 87.42%

Score 1 12.58%



# Obrigada