

# PROJETO INTEGRADOR I

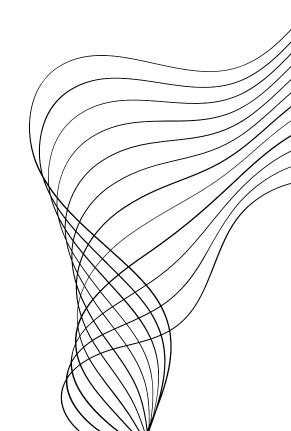
ANÁLISE DE SENTIMENTOS EM COMENTÁRIOS DE SUPERMERCADOS DE ITAPAJÉ: ENTENDENDO A OPINIÃO DOS CONSUMIDORES

**ORIENTADOR: ARTUR FRANCO** 

LARISSA VITÓRIA RUAN RODRIGUES

@LARISSAVVSOUSA@ALU.UFC.BR

@RUANRODRIGUES@ALU.UFC.BR





# QUAIS OS DADOS E SUAS APLICAÇÕES?

Os dados são avaliações e comentários referentes a três principais supermercados da cidade local. Pensamos em realizar uma análise de sentimento (NLP – processamento de linguagem natural).



# IMPACTO DO ESTUDO PARA A CIDADE

- Melhoria na qualidade dos serviços;
- Promoção da concorrência saudável;
- Aumento da satisfação do consumidor;
- Reforço da reputação da cidade;
- Atração de investimentos.



### DATASET

#### O dataset final irá conter as seguintes colunas:

- Nome do cliente;
- Comentário;
- Quantidade de curtida naquele comentário;
- Nota (estrela);
- Período;
- Nível do comentarista (local guide);

- Quantidade de fotos da pessoa no geral;
- Quantidade de comentários da pessoa no geral;
- Nome do supermercado;
- Endereço do supermercado;
- Retorno do supermercado diante da crítica/elogio.

#### Onde:

- Cada instância é um comentário de um cliente referente à um determinado supermercado;
- Possui valores ausentes.



### COLETA DE DADOS

Para a obtenção dos dados, basta realizar uma pesquisa no Google: "Supermercados em Itapajé, Ceará". A partir daí, uma lista de supermercados populares na cidade será exibida. Pensamos em utilizar apenas os principais estabelecimentos mais buscados pela comunidade local: Deirton, Para Ty e Menor Preço.









# OBTENÇÃO DOS DADOS

#### **Instant Data Scraper**

- Instant Data Scraper extrai dados de páginas da web e os exporta como arquivos Excel ou CSV;
- É uma ferramenta automatizada de extração de dados para qualquer site;
- Ele usa IA para prever quais dados são mais relevantes em uma página HTML e permite salvá-los em arquivo Excel ou CSV (XLS, XLSX, CSV).



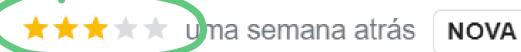


# OBTENÇÃO DAS NOTAS (JAVASCRIPT)

Coletar as notas (estrelas) de forma automatizada utilizando JavaScript

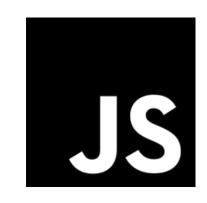


Local Guide · 17 comentários · 11 fotos



Nas compras presenciais, é o maior e melhor mercantil da cidade. No entanto, para compras online, tudo é demorado, o atendimento, o pedido, a entrega. Nesse processo, passei quase 3 horas esperando pelas compras.





# APLICAÇÃO





R. Fausto Pinheiro, 316 - Itapajé, CE



#### Assuntos frequentes

Todas preços 44 região 13 cliente 9 promoções 7 organizado 7 opções 6

#### Ordenar por

Mais relevantes Mais r

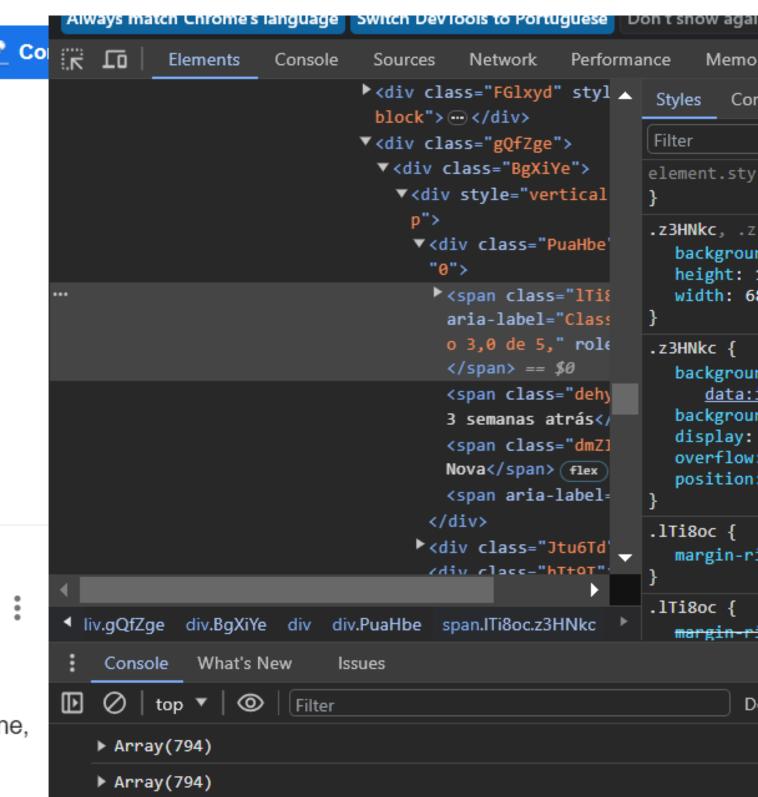
Mais recentes

Mais alta

Mais baixa



Nas compras presenciais, é o maior e melhor mercantil da cidade. No entanto, para compras online, tudo é demorado, o atendimento, o pedido, a entrega. Nesse processo, passei quase 3 horas





### 1° TENTATIVA

Falhou! Coletou notas duplicadas, com quantidade muito além do esperado/real ou muito pequena e de forma aleatoria.

```
> var res = ""
  var doc = document.querySelectorAll(".lTi8oc.z3HNkc")
  for(var i=0; i<doc.length; i++){
    res += doc[i].textContent.replace("\n", " BREAK ") + "\n"
  }
  console.log(res)</pre>
```

### 2° TENTATIVA



Sucesso! Coletou na ordem e na quantidade correta de notas.

```
> var tags = document.querySelectorAll('.c9QyIf [aria-label*="Classificado"]')
 var res = []
 for(var i=1; i<tags.length ;i+=2){</pre>
 res.push( tags[i].getAttribute('aria-label').replace("\n") + "\n" )
 console.log(res)
                                                                                               VM729:6
    (794) ['Classificado como 3,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de
    5,\n', 'Classificado como 4,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de
    5,\n', 'Classificado como 4,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de
    5,\n', 'Classificado como 3,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 4,0 de
    5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de
    5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de
    5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 3,0 de
    5,\n', 'Classificado como 4,0 de 5,\n', 'Classificado como 1,0 de 5,\n', 'Classificado como 1,0 de
    5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 3,0 de
    5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de
    5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de
    5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 4,0 de 5,\n', 'Classificado como 1,0 de
    5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de
    5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 4,0 de 5,\n', 'Classificado como 5,0 de
    5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de
    5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 4,0 de 5,\n', 'Classificado como 4,0 de
  _ 5,\n', 'Classificado como 4,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 3,0 de
    5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de
    5,\n', 'Classificado como 1,0 de 5,\n', 'Classificado como 5,0 de 5,\n', 'Classificado como 5,0 de
    5.\n'. 'Classificado como 5.0 de 5.\n'. 'Classificado como 5.0 de 5.\n'. 'Classificado como 4.0 de
```



### DETALHAMENTO DO CÓDIGO

- A primeira linha procura todos os elementos HTML que têm a classe *c9Qylf* e um atributo *aria-label* contendo a palavra "Classificado", e os resultados são armazenados na variável tags;
- Em seguida, a variável "res" é inicializada como uma lista vazia para armazenar os resultados;
- Um for é iniciado, o loop continuará enquanto i for menor que o número de elementos em tags;
- Dentro do loop, o código extrai o valor do atributo *aria-label* do elemento atual em tags[i] e adiciona à lista "res" usando push(). Isso é feito apenas para elementos em posições ímpares no array tags devido à condição i+=2 no loop;
- Por fim, o conteúdo da lista "res" é impresso no console do navegador usando console.log().

```
> var tags = document.querySelectorAll('.c9QyIf [aria-
label*="Classificado"]')
  var res = []
  for(var i=1; i<tags.length ;i+=2){
   res.push( tags[i].getAttribute('aria-label') )
  }
  console.log(res)
```



# DADOS QUE JÁ TEMOS

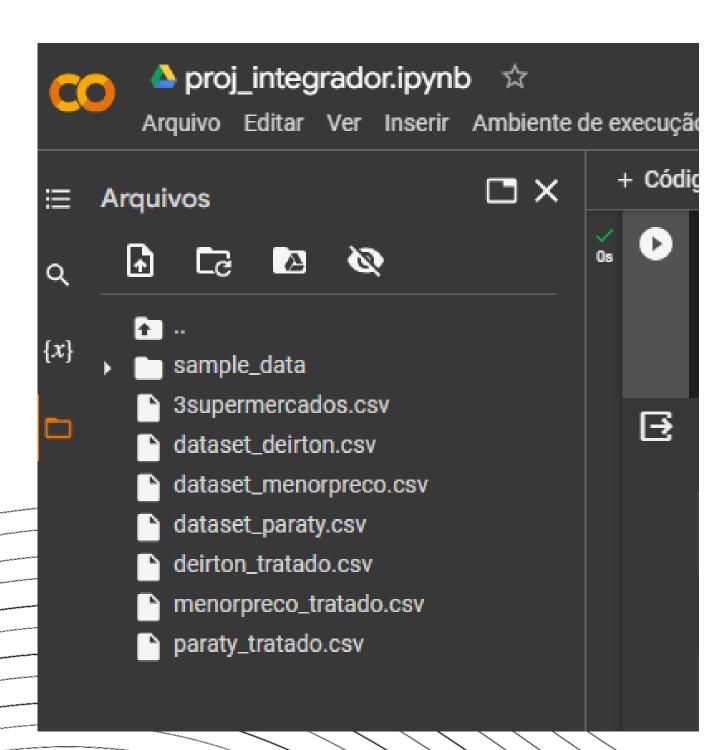
Meu Drive → Projeto Integrador I →



Tipo	▼ Pessoas ▼ Modificado ▼				
Nome	<b>↑</b>	Proprietário	Última ▼	Tamanho do	:
	dataset_deirton.csv	eu	31 de ago. de 2023	681 KB	:
	dataset_menorpreco.csv	eu	31 de ago. de 2023	160 KB	:
	dataset_paraty.csv	eu	31 de ago. de 2023	338 KB	:
	notas_deirton.csv	eu	10:07	4 KB	:
	notas_menorpreco.csv	eu	10:25	923 bytes	:
	notas_paraty.csv	eu	10:18	2 KB	:
СО	proj_integrador.ipynb 😃	eu	2 de out. de 2023	642 KB	:



# CÓDIGO DO PROJETO



```
Juntando os 3 supermercados

d = pd.read_csv("/content/deirton_tratado.csv")
mp = pd.read_csv("/content/menorpreco_tratado.csv")
p = pd.read_csv("/content/paraty_tratado.csv")
supermercados = pd.concat([d, mp, p])

supermercados.head(10)
```

#### UNIVERSIDADE FEDERAL DO CEARÁ CAMPUS DE ITAPAJÉ

## CSV COM DEZ LINHAS

								_			
	Unnamed: Ø	nome	local_guide	periodo	comentario	quant_like_coment	feedback_super	quant_comentarios	quant_fotos	supermercado	rua_supermercado
0	0	Ana Carolina Teixeira Bastos	Local Guide	3 meses atrás	O supermercado conta com produtos variados e b	0.0	Agradecemos sua ajuda.	14	17	Deirton	Fausto Pinheiro, 316
1	1	Denis sousa	Local Guide	um mês atrás	Até o momento o maior mercantil da cidade,com	0.0	Agradecemos sua ajuda.	24	4	Deirton	Fausto Pinheiro, 316
2	2	Paulo Vitor	Local Guide	3 meses atrás	Supermercado Excelente para realizar suas comp	0.0	Agradecemos sua ajuda.	34	18	Deirton	Fausto Pinheiro, 316
3	3	Karla Patricia Dos Santos Pinto	Not Local Guide	4 meses atrás	Atendimento maravilhoso, os funcionários são p	0.0	Agradecemos sua ajuda.	2	0	Deirton	Fausto Pinheiro, 316
4	4	Ricardo Nascimento	Local Guide	um mês atrás	Ótimo atendimento. Promoções e variedade	0.0	Agradecemos sua ajuda.	6	16	Deirton	Fausto Pinheiro, 316
356	356	ZARIFF LUTIF PESSOA ROCHA	Not Local Guide	3 anos atrás	Sem comentário	0.0	Sem feedback	0	0	Paraty	Bacharel José Airtor Rocha, 1655
357	357	xavier mota	Not Local Guide	3 anos atrás	Sem comentário	0.0	Sem feedback	0	19	Paraty	Bacharel José Airtor Rocha, 1655
358	358	Mônica Vieira	Local Guide	4 anos atrás	Sem comentário	0.0	Sem feedback	25	7	Paraty	Bacharel José Airton Rocha, 1655
359	359	Luciana	Not Local Guide	2 anos atrás	Sem comentário	0.0	Sem feedback	0	0	Paraty	Bacharel José Airton Rocha, 1655
360	360	rurik alves	Not Local Guide	um ano atrás	Sem comentário	0.0	Sem feedback	8	0	Paraty	Bacharel José Airton Rocha, 1655
1331	rows × 11 colu	ımns									



# PRÓXIMOS PASSOS

- Dados com as notas (estrelas). 🕥
- Pré-tratamento: troquem os nomes por números. 🏈
- Pré-tratamento de texto: remoção de stopwords (preposições, artigos etc), tokenização. (NLTK ou SpaCy)
- (Opcional) Análise de emojis.



## ATUALIZAÇÃO DOS DADOS

Meu Drive > Projeto Integrador I -**(i)**  $\blacksquare$ Modificado ▼ Tipo ▼ Pessoas 🔻 Última ... ▼ Proprietário Tamanho do Nome deirton\_bd.csv eu 23 de out. de 2023 97 KB menorpreco\_bd.csv 23 de out. de 2023 22 KB eu notas\_d.csv 23 de out. de 2023 2 KB eu eu 23 de out. de 2023 559 bytes notas\_mp.csv 23 de out. de 2023 notas\_p.csv eu 1 KB paraty\_bd.csv 23 de out. de 2023 44 KB eu proj\_integrador.ipynb 🚢 627 KB 09:40 eu

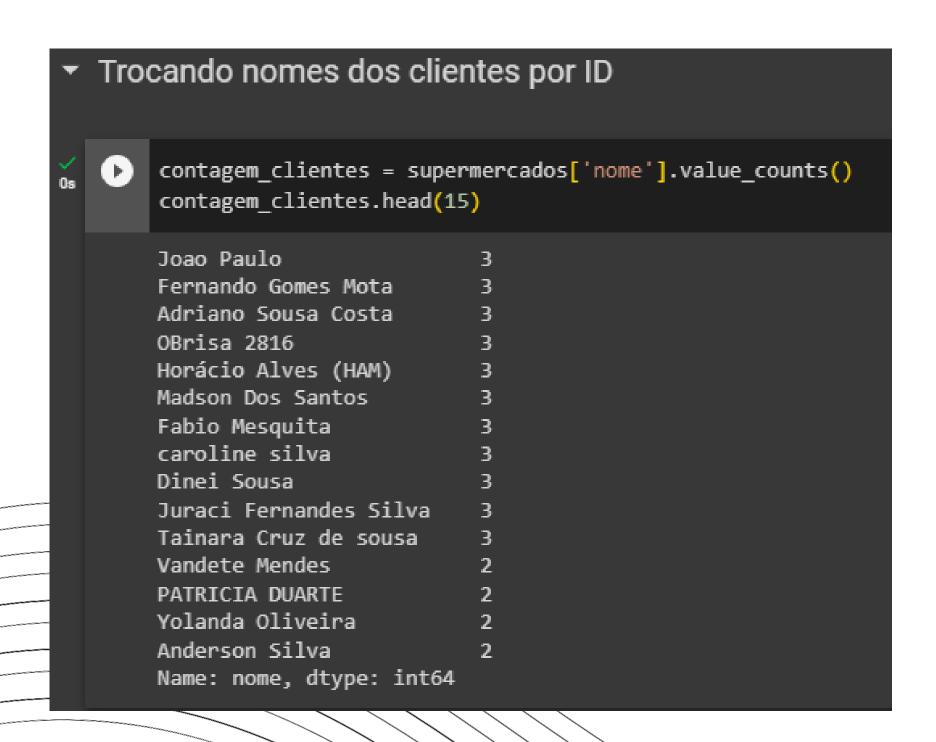
#### UNIVERSIDADE FEDERAL DO CEARÁ CAMPUS DE ITAPAJÉ

### DATASET COM NOTAS

	nome	periodo	comentario	quant_like_coment	feedback super	nota	quant_comentarios	quant fotos	supermercado	rua_supermercado	guia local	cliente id
	Hone	2 anos	Concilcul 10	quarre_11ke_comerre	recuback_super	nocu	quant_concilcul 103	quant_rocos	Super mer educ	Tua_super mer cauo	guiu_iocui	CIICHCC_IU
772	carlos Rocha	atrás	Sem comentário	0.0	Sem feedback	3	2	2	Deirton	Fausto Pinheiro, 316	Não	767
258	Eliene Rocha	3 anos atrás	Sem comentário	0.0	Sem feedback	5	0	0	Paraty	Bacharel José Airton Rocha, 1655	Sim	463
111	laercio lima	2 anos atrás	Ótimo lugar pra fazer suas compras	0.0	Agradecemos sua ajuda.	5	6	0	Deirton	Fausto Pinheiro, 316	Não	111
503	helsing loureto	3 anos atrás	Sem comentário	0.0	Sem feedback	4	58	107	Deirton	Fausto Pinheiro, 316	Sim	501
303	ZARIFF LUTIF PESSOA ROCHA	3 anos atrás	Sem comentário	0.0	Sem feedback	5	0	0	Paraty	Bacharel José Airton Rocha, 1655	Não	1047
73	Rany Sousa	um ano atrás	Sem comentário	0.0	Sem feedback	4	5	0	Paraty	Bacharel José Airton Rocha, 1655	Sim	918
159	André Luiz	5 anos atrás	Comodidade e conforto a um preço acessível.	0.0	Agradecemos sua ajuda.	5	3	0	Deirton	Fausto Pinheiro, 316	Não	159
549	Aline Bastos	5 anos atrás	Sem comentário	0.0	Sem feedback	5	0	0	Deirton	Fausto Pinheiro, 316	Não	547
156	Vanderli Barbosa Pinto	4 anos atrás	Sem comentário	0.0	Sem feedback	5	0	0	Menor preço	Manoel Luís da Rocha, 87	Não	872
201	Rafael Miguel	10 meses atrás	Ótima	0.0	Agradecemos sua ajuda.	1	3	0	Deirton	Fausto Pinheiro, 316	Não	201
60	daniele patricio	um ano atrás	Sem comentário	0.0	Sem feedback	5	0	34	Paraty	Bacharel José Airton Rocha, 1655	Não	912
272	Joyce Costa	3 anos atrás	Sem comentário	0.0	Sem feedback	4	14	0	Paraty	Bacharel José Airton Rocha, 1655	Não	181



### TROCANDO NOME POR ID





1344 rows × 11 columns

595 rows × 11 columns



### TROCANDO NOME POR ID

```
unique_clientes = supermercados['nome'].unique()
cliente_id_mapping = {client: idx for idx, client in enumerate(unique_clientes)}
supermercados['cliente_id'] = supermercados['nome'].map(cliente_id_mapping)
```



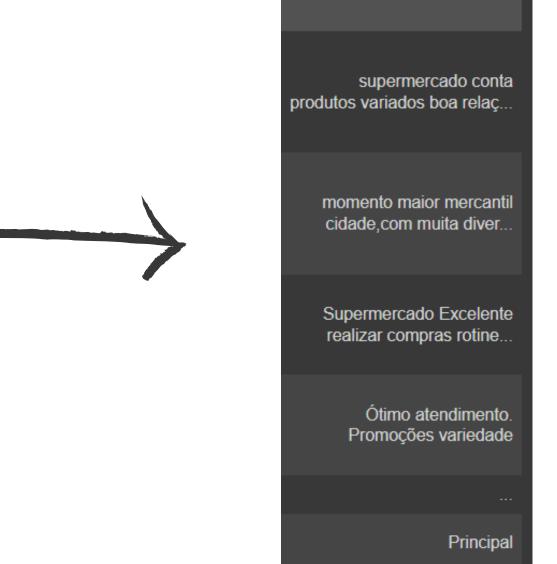
cliente_i	d	
	0	
	1	
	2	
	3	
	4	
30	6	
41	1	
50	00	
87	2	
79	)5	



### REMOVENDO STOPWORD's

(preposições, artigos etc)

Stopwords são palavras que podem ser consideradas irrelevantes para o entedimento do sentido de um texto, ou seja, palavras semanticamente irrelavantes. Exemplos: as, e, os, de, para, com, sem, foi.



comentario\_sem\_stopwords

compras presenciais, maior

melhor mercantil ci..

Sem comentário



### REMOVENDO STOPWORD'S

#### Antes

#### comentario

Nas compras presenciais, é o maior e melhor me...

supermercado conta com produtos variados e b...

Até o momento o maior mercantil da cidade,com ...

#### Depois

comentario\_sem\_stopwords

compras presenciais, maior melhor mercantil ci...

supermercado conta produtos variados boa relaç...

momento maior mercantil cidade,com muita diver...



# PRÓXIMOS PASSOS

- Remoção de pontuação
- Pré-tratamento de texto: tokenização (NLTK ou SpaCy) 🏈
- Ranking de 20 palavras 🏈
- Distribuição das notas
- Nuvem de palavras





# REMOVENDO PONTUAÇÃO

Antes

comentario\_sem\_stopwords

compras presenciais, maior melhor mercantil ci...

supermercado conta produtos variados boa relaç...

momento maior mercantil cidade,com muita diver...

Supermercado Excelente realizar compras rotine...

Ótimo atendimento. Promoções variedade Depois

comentario\_sem\_pontuacao

compras presenciais maior melhor mercantil cid...

supermercado conta produtos variados boa relaç...

momento maior mercantil cidadecom muita divers...

Supermercado Excelente realizar compras rotine...

Ótimo atendimento Promoções variedade



# TOKENIZAÇÃO

Depois Antes

comentario\_sem\_pontuacao

compras presenciais maior melhor mercantil cid...

supermercado conta produtos variados boa relaç...

> momento maior mercantil cidadecom muita divers..

> Supermercado Excelente realizar compras rotine..

Ótimo atendimento Promoções variedade

comentario\_tokens

[compras, presenciais, maior, melhor, mercanti.

> [supermercado, conta, produtos, variados, boa,

[momento, maior, mercantil. cidadecom, muita,

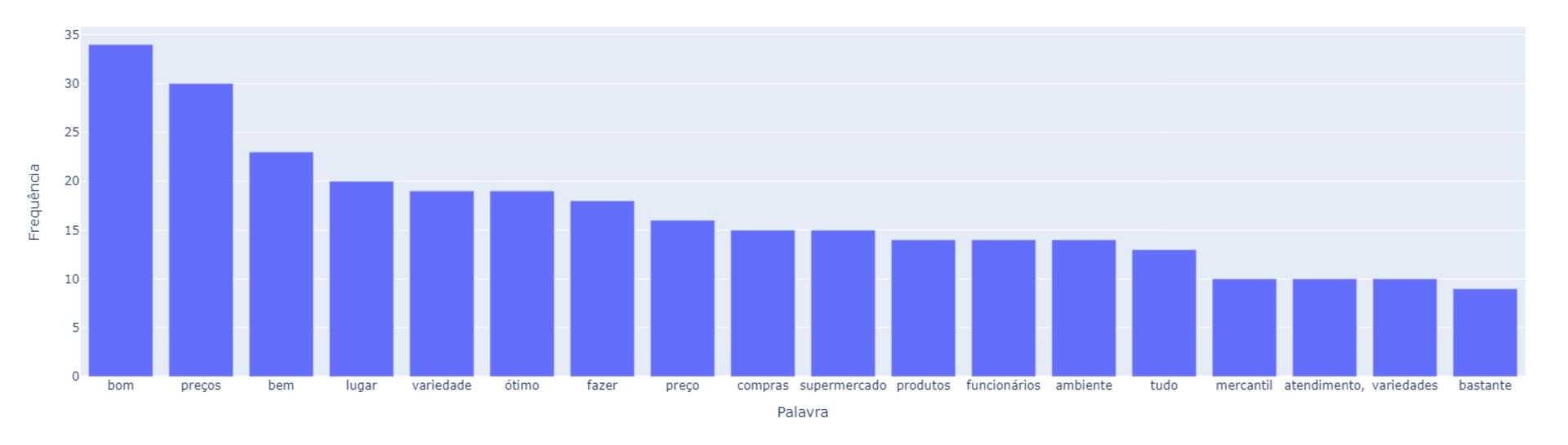
[Supermercado, Excelente, realizar, compras, r.

[Ótimo, atendimento, Promoções variedade]



### RANKING 20 PALAVRAS

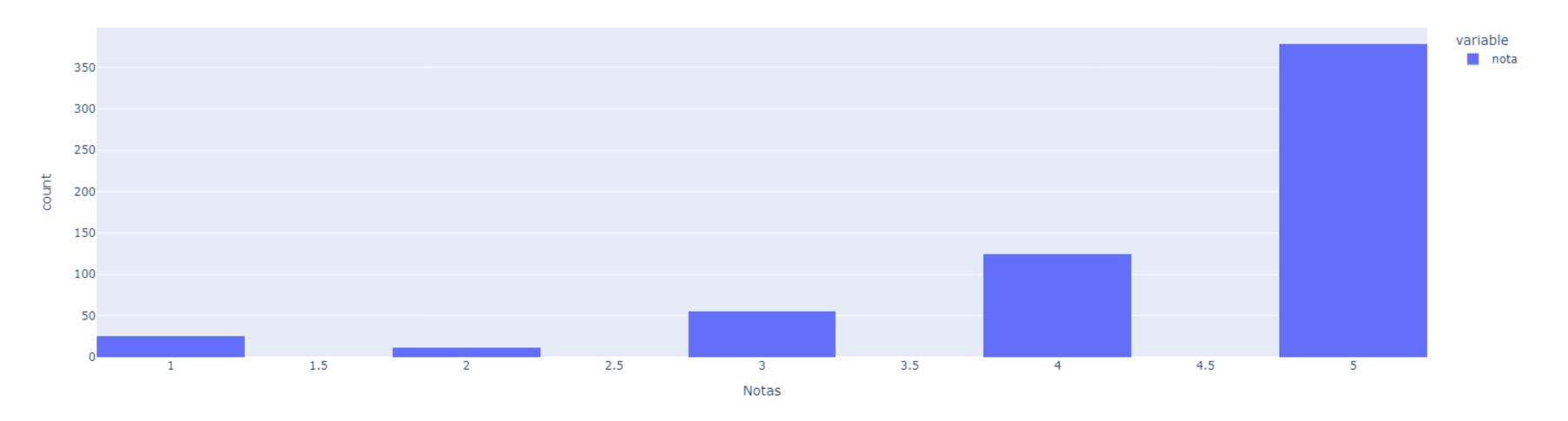
Top 20 Palavras Mais Frequentes





# DISTRIBUIÇÃO DAS NOTAS

#### Histograma das Notas





### NUVEM DE PALAVRAS







#### Noções básicas de JavaScript

Disponível em: https://developer.mozilla.org/pt-BR/docs/Learn/Getting\_started\_with\_the\_web/JavaScript\_basics

Análise de Sentimentos de Reviews de Produtos em E-commerces Brasileiros Autor(es): Carmo Neto, Francisco Vieira

Análise de sentimentos das reviews da Olist

Autor(es): Amanda Alves

Análise de Sentimento para Reviews Apresentados em Vídeos: Modelo de Redes Neurais Treinado em Base de Reviews Escritos

Autor(es): João Paulo, Romulo Baldez, Caio César, Raquel Cristina, Cristiano Gonçalves, Gustavo Correa

