

## 1. Pré-Processamento de Dados

No script, foi necessário a realização de alguns processos de pré-processamento que não somente facilitam a mineração de dados para ulterior análise, como também a tornam possível. Dessa forma, os processos visam transformar os dados brutos extraídos do arquivo JSON `"padaria_trab.json"` em matrizes manipuláveis para a análise de padrões de associação de produtos.

### 1.1 Criação das Listas `header` e `header_reduced`:

- **Objetivo:** Agrupar variações de produtos em categorias principais para identificar tendências de compra não só entre produtos específicos, mas também em categorias gerais, como "Pão" e "Queijo".
- **Explicação:** `header` foi usado para capturar cada produto individual, enquanto `header_reduced` foi projetado para categorizar produtos em tipos principais, atendendo ao requisito de análise de associações em um nível mais detalhado.

### 1.2 Construção das Matrizes `df` e `df_reduced`:

- **Objetivo:** Criar representações binárias das transações para execução eficiente do algoritmo Apriori.
- **Explicação:** `df` permite a análise de produtos específicos, enquanto `df_reduced` possibilita analisar padrões mais amplos por categoria. Isso é essencial para identificar associações frequentes e compreender a compra conjunta de categorias de produtos.

### 1.3 Listagem de Doces (`doces` e `not_doces`) e Matriz `mat_doces`:

- **Objetivo:** Focar na identificação de associações que impliquem a compra de doces, um dos principais objetivos da análise.
- **Explicação:** A criação de listas específicas para doces e não-doces facilita a extração de regras direcionadas, como solicitado na segunda parte do trabalho. `mat_doces` foi estruturada para capturar compras de doces em relação a outros produtos, facilitando a filtragem no algoritmo Apriori.

### 1.4 Combinação de Matrizes para Associação com Doces (`df_doces`):

- **Objetivo:** Garantir que as regras com doces no consequente sejam extraídas com a estrutura adequada.
- **Explicação:** `df_doces` integra todos os produtos que não são doces e a coluna "Doce" para que o Apriori possa gerar regras focadas na associação com a compra de doces, o que atende diretamente à necessidade de encontrar padrões do tipo `{Produto} => {Doce}`

## 2. Execução e Resultado do Algoritmo Apriori

O código utiliza o algoritmo Apriori para identificar regras de associação nos dados. Foram criados três conjuntos de regras:

### 2.1 Regras Gerais (**rules**):

- Aplicado na matriz completa **df**, com suporte mínimo (**supp**) de 5% e confiança mínima (**conf**) de 40%.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Presunto Perdigão}	=> {Pão Francês}	0.06504065	0.5000000	0.1300813	2.196429	8
[2]	{Presunto Perdigão}	=> {Queijo Mussarela}	0.05691057	0.4375000	0.1300813	1.379808	7
[3]	{Presunto Sadia}	=> {Queijo Mussarela}	0.06504065	0.4000000	0.1626016	1.261538	8
[4]	{Pão Francês}	=> {Queijo Mussarela}	0.13008130	0.5714286	0.2276423	1.802198	16
[5]	{Queijo Mussarela}	=> {Pão Francês}	0.13008130	0.4102564	0.3170732	1.802198	16

### 2.2 Regras Reduzidas (**rules\_reduced**):

- Executado na matriz **df\_reduced**, suporte mínimo de 10% e confiança mínima de 58%. Aqui, são analisadas associações em um nível mais genérico, agrupando os produtos pelo seu tipo principal.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Presunto}	=> {Pão}	0.2520325	0.5849057	0.4308943	1.284704	31
[2]	{Pão}	=> {Queijo}	0.2845528	0.6250000	0.4552846	1.260246	35
[3]	{Pão, Presunto}	=> {Queijo}	0.1707317	0.6774194	0.2520325	1.365944	21
[4]	{Queijo, Presunto}	=> {Pão}	0.1707317	0.7000000	0.2439024	1.537500	21
[5]	{Queijo, Pão}	=> {Presunto}	0.1707317	0.6000000	0.2845528	1.392453	21

### 2.3 Regras com Doces no Consequente (**rules\_doce**):

- Regras específicas em que o item (**rhs**) é "Doce". O conjunto de regras utiliza a matriz **df\_doces**, suporte mínimo de 5% e confiança mínima de 40%.

[1]	{Queijo Minas}	=> {Doce}	0.05691057	0.5384615	0.1056911	1.439799	7
[2]	{Pão Gajeta}	=> {Doce}	0.05691057	0.5000000	0.1138211	1.336957	7
[3]	{Pastel Frango}	=> {Doce}	0.06504065	0.4705882	0.1382114	1.258312	8
[4]	{Refri - Fanta}	=> {Doce}	0.08130081	0.5882353	0.1382114	1.572890	10
[5]	{Café Melita}	=> {Doce}	0.07317073	0.5294118	0.1382114	1.415601	9

## 3. Respostas e Análise das Regras

Baseando-se nas regras geradas para cada uma das matrizes, temos os seguintes insights:

### 3.1 Análise das Regras Gerais:

As regras obtidas com todos os produtos evidenciam que a regra **{Pão Francês} => {Queijo Mussarela}** é a mais influente. Isso se deve pois a confiança é de 57.1%, ou seja, mais da metade das compras que têm como item pão francês, têm queijo mussarela, e o lift é de 1.80, que significa que a compra de pão francês aumenta significativamente a compra de queijo mussarela. Além disso, a regra **{Presunto Perdigão} => {Pão Francês}** também é influente, pois a compra de presunto da marca Perdigão aumenta significativamente a compra de pão francês.

### 3.2 Análise das Regras Reduzidas:

As regras obtidas apenas com as categorias dos produtos mostram padrões mais amplos e indicam que certos tipos de produtos, como pães, queijo e presunto, frequentemente aparecem juntos. A mais influente destas regras é **{Queijo, Presunto} => {Pão}**, pois a sua confiança é de 70%, portanto 70% das compras de queijo e presunto possuem pão, seu lift é de 1.54 e seu suporte é de 17.1%, mostrando que é uma combinação de compras frequente.

### 3.3 Análise das Regras com Doces no Consequente:

O conjunto que contém as associações entre os produtos e a categoria “Doce” evidencia que entre os produtos analisados, o Refri - Fanta é o item que mais implica a compra de doces, devido ao maior lift, suporte e a confiança mais alta. Isso significa que os consumidores identificam que o doce é um bom complemento para refrigerantes, em especial o refrigerante Fanta. Além disso, as outras regras evidenciam que os doces frequentemente são comprados como complemento ou sobremesa de lanches rápidos e até bebidas, como café.

## Conclusão do Documento

Com base nas regras apresentadas, podemos concluir que existe uma forte relação de consumo entre produtos alimentícios complementares, como presunto, pão e queijo, que frequentemente aparecem juntos nas transações. Por exemplo, a associação entre presunto e pão, e pão e queijo, possuem lift e confiança altos, o que significa que os consumidores consideram esses produtos como básicos e essenciais para consumo cotidiano muitas vezes sendo comprados em conjunto.

Além disso, observamos uma tendência de compra de produtos doces acompanhando lanches variados, como queijo minas, pão gajeta e até bebidas como refrigerante Fanta. A presença de confiança significativa nas combinações com doces reforça a ideia de que doces são frequentemente adquiridos como complemento ou sobremesa.

Essas informações podem ser utilizadas para estratégias de disposição e promoções de produtos, visando aumentar as vendas de produtos considerados complementares.