

UNIVERSIDADE FEDERAL DE SANTA MARIA CENTRO DE TECNOLOGIA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Arthur Bogacki Verissimo e Leandro Oliveira G. do Nascimento

TRABALHO 1: PADARIA MINERAÇÃO

Santa Maria, RS
2024

RESUMO

TRABALHO 1: PADARIA MINERAÇÃO

AUTORES: Arthur Bogacki Verissimo e Leandro Oliveira G. do Nascimento

PROFESSOR: Joaquim V C Assunção

Este relatório tem o objetivo de descrever os passos realizados na mineração de dados de uma padaria. O objetivo principal é realizar o devido pré-processamento de um arquivo **padaria_trab.json**, extrair as principais regras, encontrar o produto mais influente 1 para 1 e encontrar a regra dos produtos que implicam compra de “doce”.

Relatório:

Pré-Processamento:

O projeto "Padaria Mineração" tem como objetivo analisar as associações entre produtos vendidos em uma padaria. Para isso, focamos em identificar quais produtos são frequentemente adquiridos juntos. Os principais produtos em análise incluem: Café, Pão, Presunto, Queijo, Pastel, Doce e Refrigerante.

Etapas do Processo

1 - Carregamento do Arquivo JSON

A primeira etapa foi carregar o arquivo JSON que contém os dados de vendas. Para transformar esses dados em um formato utilizável, utilizamos a função `json_normalize` do Python, que converte o arquivo JSON em um DataFrame, facilitando a manipulação e análise dos dados.

2 - Remoção de Colunas Irrelevantes

No segundo passo, removemos a coluna "compra", que continha um número igual para todas as entradas. Essa coluna não contribuía para a análise de associação e, portanto, sua remoção foi crucial para focar nos produtos de interesse.

3 - Formatação dos Nomes

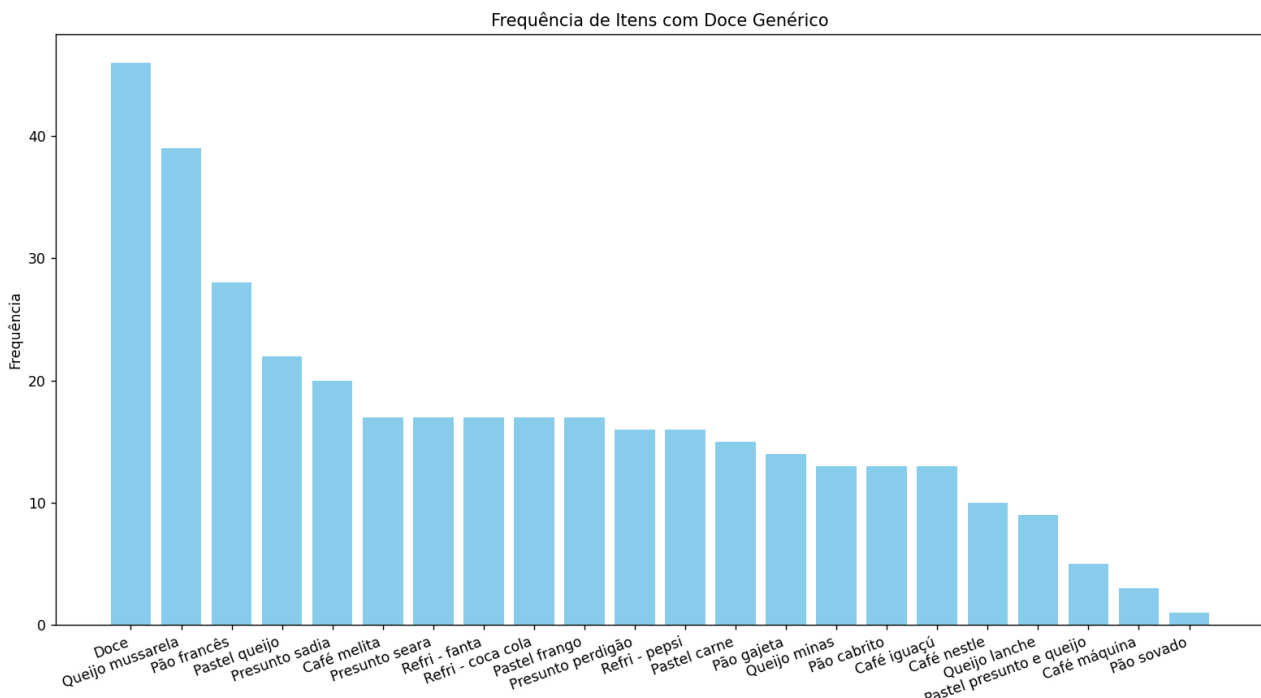
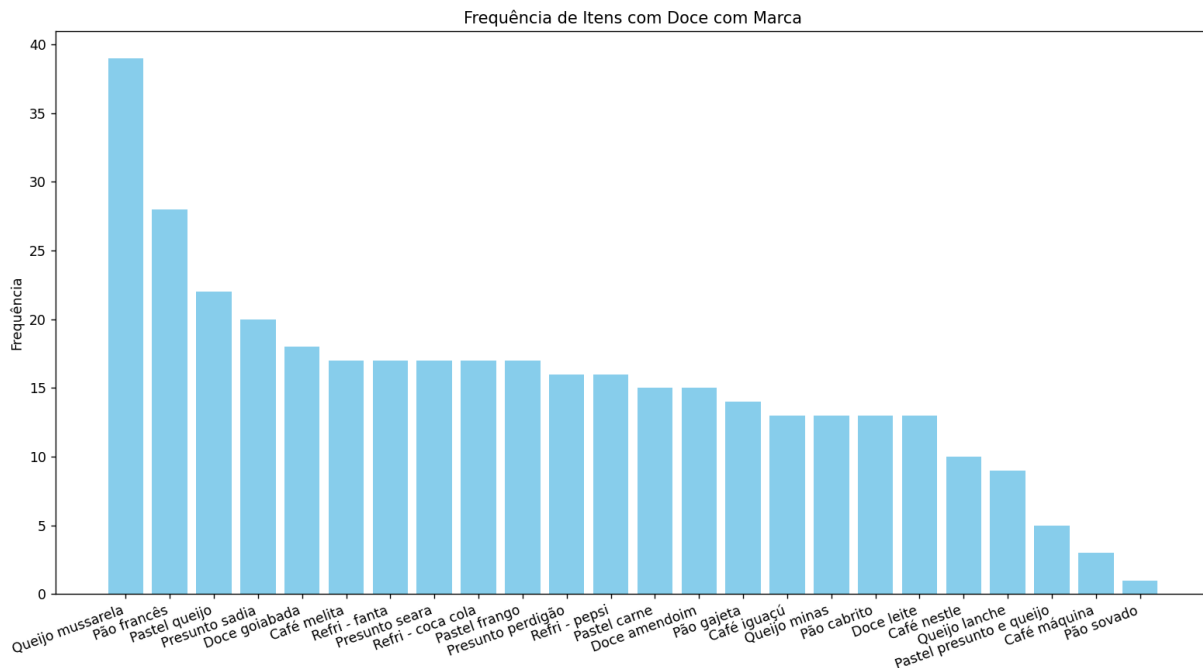
Por fim, formatamos os nomes dos produtos para garantir uniformidade. Essa etapa é importante para evitar discrepâncias nas nomenclaturas e facilitar a análise, assegurando que cada produto seja reconhecido da mesma forma, independentemente de variações em sua escrita. Nesta etapa precisamos criar outra coluna no data-frame, um com Doce "Genérico" e outro com o Doce e suas respectivas marcas. Assim o `df['produtos']` ficou com doces com suas marcas e `df['produtos2']` ficou apenas doces.

4 - Transformação em Colunas de True e False

Nessa etapa, criamos uma coluna para cada produto, indicando True se o produto está presente na compra e False se não está. Isso foi feito através da explosão da coluna de produtos e da aplicação do one-hot encoding, permitindo que cada produto fosse representado como uma coluna binária. Após agrupar os dados, as ocorrências foram somadas e convertidas para valores booleanos.

5. Análise das ocorrências

Utilizamos um gráfico criado pela biblioteca do python matplotlib para analisar as ocorrências e verificar se o que aplicamos nas etapas anteriores estão corretas.



6 - Aplicar o Princípio Apriori

Utilizamos o Princípio Apriori para identificar conjuntos frequentes de itens. Esse princípio nos permite encontrar quais produtos são frequentemente comprados juntos e gerar regras de associação baseadas em métricas como suporte, confiança e lift.

7 - Análise das Regras

As 5 principais regras de associação foram exibidas, juntamente com a regra mais influente que mostra a relação direta entre dois produtos. Além disso, foram identificadas as regras que implicam a compra de "Doce", permitindo uma visão mais detalhada sobre as preferências dos consumidores.

8 - Interpretação das Regras

As regras geradas fornecem informações valiosas sobre como os produtos são comprados juntos. O antecedente é o produto analisado, como "presunto", enquanto o conseqüente é o produto frequentemente adquirido junto, como "pão francês". O suporte indica a frequência dessa regra em relação ao total de transações. Por exemplo, um suporte de 0,1 significa que a combinação foi observada em 10% das compras. A confiança mostra a probabilidade de um cliente comprar o conseqüente após adquirir o antecedente; uma confiança de 0,8 indica que, em 80% das vezes que alguém compra presunto, também compra pão francês. Por fim, o lift compara a frequência do conseqüente quando o antecedente é comprado com sua frequência geral. Um lift maior que 1 sugere que a compra do antecedente aumenta as chances de adquirir o conseqüente.

9 - Resultados da Mineração

Regras de Associação:

Total de regras de associação encontradas: 6

antecedents	consequents	support	confidence	lift
(Presunto perdigão)	(Pão francês)	0.065041	0.500000	2.196429
(Pão francês)	(Presunto perdigão)	0.065041	0.285714	2.196429
(Presunto sadia)	(Queijo mussarela)	0.065041	0.400000	1.261538
(Queijo mussarela)	(Presunto sadia)	0.065041	0.205128	1.261538
(Pão francês)	(Queijo mussarela)	0.130081	0.571429	1.802198
(Queijo mussarela)	(Pão francês)	0.130081	0.410256	1.802198

As 5 principais regras de associação:

antecedents	consequents	support	confidence	lift
(Presunto perdigão)	(Pão francês)	0.065041	0.500000	2.196429
(Pão francês)	(Presunto perdigão)	0.065041	0.285714	2.196429
(Queijo mussarela)	(Pão francês)	0.130081	0.410256	1.802198
(Pão francês)	(Queijo mussarela)	0.130081	0.571429	1.802198
(Queijo mussarela)	(Presunto sadia)	0.065041	0.205128	1.261538

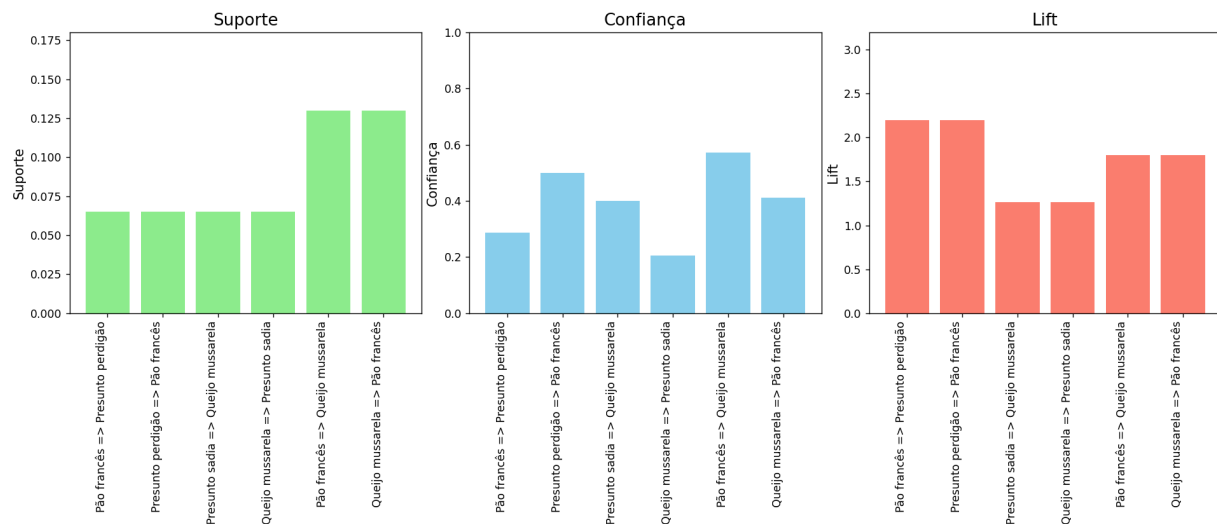
A regra mais influente (produto 1 para 1):
['Presunto perdigão'] => ['Pão francês']

```

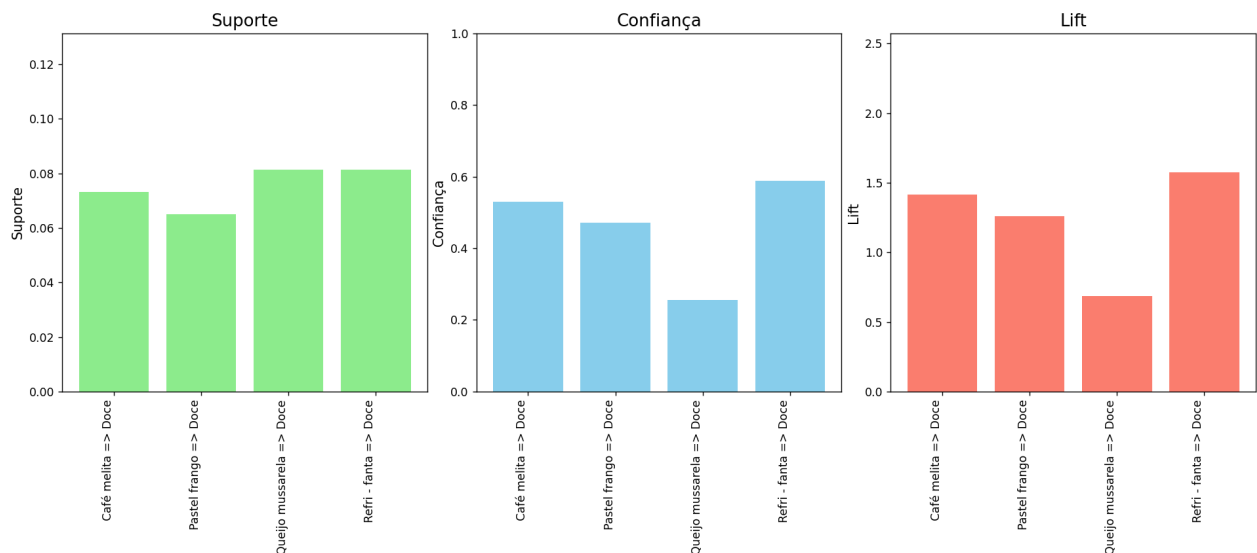
Regras que implicam a compra de 'Doce':
  antecedents consequents support confidence lift
(Café melita)      (Doce) 0.073171 0.529412 1.415601
(Pastel frango)    (Doce) 0.065041 0.470588 1.258312
(Queijo mussarela) (Doce) 0.081301 0.256410 0.685619
(Refri - fanta)    (Doce) 0.081301 0.588235 1.572890

```

Todas Regras de Associação



Regras que implicam Doce



Para a extração das regras foi utilizado o suporte mínimo de 0.06

```

# Principio Apriori:
frequent_itemsets = apriori(df_final, min_support=0.06, use_colnames=True)

```

10. Conclusão

O que fazer com esses dados?

1. Promoções Combinadas

- "Presunto Perdigão" e "Pão Francês": Essa é a regra mais forte, com um *lift* de 2.19, indicando que clientes que compram um tendem a comprar o outro. Uma promoção combinada desses itens (por exemplo, desconto ao comprar os dois juntos) pode incentivar ainda mais as vendas.
- "Pão Francês" e "Queijo Mussarela": Também apresentam associação significativa. A padaria pode promover essas combinações como um "kit lanche", oferecendo uma vantagem adicional para a compra conjunta.

2. Planejamento de Layout de Loja

- Agrupamento de Produtos Relacionados: Organizando itens como presunto, pão francês e queijo próximo um do outro, facilitará o acesso e promoverá mais a compra por impulso.
- Doce como Complemento de Compras: Itens como "Café Melita" e "Doce" têm uma associação forte. Expor doces perto do café pode aumentar a probabilidade de venda conjunta.

3. Personalização de Ofertas

- Descontos em Produtos Complementares: Com base em itens que o cliente já comprou (como presunto ou queijo), a padaria poderia oferecer um desconto no produto associado (como pão ou doce).
- Ofertas para Aumentar Confiança Menor: Para produtos com uma confiança de associação mais baixa, como "Queijo Mussarela" e "Doce", a criação de ofertas específicas pode ajudar a fortalecer essa associação, incentivando os clientes a experimentar a combinação.

4. Gestão de Estoque e Planejamento de Compras

- Repor Estoque de Itens Associados: Com uma visão clara de quais itens são frequentemente comprados juntos, a padaria pode otimizar sua gestão de estoque, garantindo que os produtos associados estejam sempre disponíveis. Isso não apenas melhora a satisfação do cliente, mas também minimiza as perdas de vendas devido à falta de estoque.

Essas ações não apenas melhoram a experiência do cliente, mas também ajudam a otimizar as estratégias de vendas. Ao aproveitar os padrões identificados pela análise de regras de associação, a padaria pode maximizar suas vendas, melhorar a eficiência operacional e fidelizar os clientes.