

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo Campus Birigui

Bacharelado em Engenharia da Computação

Documento de Ciências de Dados

Vendas de Calçados

| Discentes: | Matrículas: |
| --- | --- |
| Nicolas Fagnani | BI3024474 |
| Gustavo Amaral Azevedo | BI3025055 |
| Marcus Vinicius Magri | BI3023907 |
| Miguel Totti de Oliveira | BI3023613 |
| Pedro Henrique Tamburi | BI3024555 |

Docente: Cleverson Moreira de Souza Turma: 231

Componente Curricular: Ciência de Dados

Birigui – São Paulo 2024

Sumário

[1.](#_zcolyrc4f538) Introdução 3

[2.](#_fvkov7v052n0) Processo de Mineração de Dados 3

[3.](#_mla4ydciyz83) Considerações Finais 3

# Introdução

O presente relatório descreve o processo de mineração de dados realizado sobre um conjunto de dados de vendas de calçados, contendo 198.410 registros, que incluem o nome do produto vendido, o cliente, a cidade e o mês da transação. O objetivo principal é identificar padrões de compra entre os produtos, utilizando o algoritmo Apriori para extrair as regras de associação.

# Processo de Mineração de Dados

### Importação das Bibliotecas

Nesta etapa, são carregadas bibliotecas essenciais para a análise:

* pandas: para manipulação de dados.
* TransactionEncoder: para transformar os dados em um formato binário apropriado para o algoritmo Apriori.
* apriori e association\_rules do mlxtend.frequent\_patterns: responsáveis por extrair os conjuntos frequentes e gerar as regras de associação.

### Leitura dos Dados

O arquivo CSV com os dados de vendas é carregado em um DataFrame. Essa base contém informações de compras realizadas, no formato onde cada linha representa uma transação com os itens comprados.

### Pré-processamento dos Dados

Essa etapa prepara os dados para serem analisados pelo Apriori. As principais ações incluem:

* Conversão das transações em listas de itens.
* Uso do TransactionEncoder para transformar essa lista em uma matriz binária (1 para presença do item e 0 para ausência), resultando em um DataFrame com valores booleanos.

Isso é necessário porque o algoritmo Apriori trabalha com transações que indicam claramente se um item está presente ou não.

### Execução do Algoritmo Apriori

Aqui é aplicado o algoritmo propriamente dito:

* O algoritmo Apriori é executado sobre a matriz binária com um valor mínimo de suporte definido em 1% ou 0.01.
* O resultado é uma tabela com os conjuntos frequentes, ou seja, combinações de itens que ocorrem com frequência suficiente nas transações.

Cada linha dessa tabela inclui o conjunto de itens e seu respectivo suporte (a proporção de transações em que o conjunto aparece).

### Geração das Regras de Associação

Depois de identificar os conjuntos frequentes, o passo seguinte é extrair regras de associação, que representam relações do tipo: "Se o cliente comprou X, então provavelmente também comprou Y".

* A função association\_rules é usada para gerar essas regras.
* São especificados critérios como confiança mínima de 85% ou 0.85 e lift, que mede a força da associação em comparação com a aleatoriedade.

O resultado é uma tabela contendo:

* Antecedentes (parte "se" da regra).
* Consequentes (parte "então").
* Suporte.
* Confiança.
* Lift.

### Filtragem das Regras

As regras geradas são filtradas para selecionar apenas aquelas com métricas mais relevantes, normalmente:

* Confiança superior a 85%.
* Lift significativamente maior que 1 (indicando dependência positiva entre os itens).

Essa filtragem ajuda a destacar as regras mais úteis e significativas para interpretação.

Depois da filtragem padrão do apriori, foi feita outra filtragem com base na métrica suporte e na quantidade de itens antecedentes em cada regra:

* Suporte superior a 2% ou 0.02.
* Quantidade de itens antecedentes inferior ou igual a 2.

Esse filtro ajudou a selecionar melhor as regras que acreditamos serem mais relevantes, com lift e confiança altos indicando que as regras são fortes e podem ser muito úteis para planejamentos posteriores.

# Considerações Finais

1. Principais Regras Geradas

As regras extraídas envolvem majoritariamente calçados da linha “KLIN” para recém-nascidos, com combinações entre sandálias, mocassins, sapatilhas e tênis. Exemplos de regras:

**Se um cliente compra**:

* *CALÇADO KLIN MOCASSIM MASCULINO RECÉM NASCIDO* e *CALÇADO KLIN SANDÁLIA FEMININA RECÉM NASCIDO* → então **também compra** *CALÇADO KLIN SANDÁLIA MASCULINA RECÉM NASCIDO*
* **Confiança:** 90,4%
* **Lift:** 14,6
* **Suporte:** 2,87%

Outra regra:

* *CALÇADO KLIN TÊNIS MASCULINO RECÉM NASCIDO* e *CALÇADO KLIN SANDÁLIA FEMININA RECÉM NASCIDO* → então **também compra** *CALÇADO KLIN SAPATILHA FEMININA RECEM NASCIDO*
* **Confiança:** 91,5%
* **Lift:** 12,05
* **Suporte:** 3

1. Métricas Relevantes das Regras

* **Confiança (85% a 91%)**: As regras apresentam altíssima confiabilidade, indicando que o consequente quase sempre aparece junto ao antecedente.
* **Lift (de 11 a 14)**: Todos os valores de lift estão muito acima de 1, o que revela **forte correlação positiva** entre os produtos — ou seja, as compras não são aleatórias, e sim associadas.
* **Representatividade = 1.0**: Todas as regras obtidas são perfeitamente representativas dentro dos conjuntos frequentes, o que sugere regras muito coesas e consistentes.
* **Jaccard e Kulczynski**: Esses valores indicam interseção relevante e simetria entre os conjuntos, confirmando a força das regras.

1. Conclusões Detalhadas

* **Padrões Fortes de Co-compra entre Gêneros e Tipos**:  
   Clientes que compram sandálias femininas para recém-nascidos frequentemente adquirem também sapatos masculinos (e vice-versa), indicando um **perfil de compra que atende múltiplos gêneros**, possivelmente irmãos ou um enxoval completo.
* **Produtos Complementares ou Substitutos Suaves**:  
   A relação forte entre tênis, sapatilhas e mocassins sugere que os clientes veem esses itens como **complementares** ou **variantes substituíveis**, mas frequentemente comprados juntos — talvez por preferência de estilo, cor ou ocasiões diferentes.
* **Alto Potencial de Cross-selling**:  
   As regras obtidas podem ser utilizadas em sistemas de recomendação (“Clientes que compraram isso também compraram...”), especialmente úteis para plataformas de e-commerce infantil.
* **Mix de Produtos Associados**:  
   A associação entre sandália, tênis e sapatilha feminina demonstra um **kit de produtos desejados por compradores**, podendo indicar bundles (combos) ideais a serem oferecidos.
* **Força Estatística Real**:  
   Com lift muito alto e suporte relevante (> 2,8%), as regras não são apenas confiáveis mas também ocorrem com frequência suficiente para **serem relevantes no negócio**.

1. **Sugestões Estratégicas com Base nos Resultados**

* **Campanhas Promocionais conjuntas**: Oferecer descontos em sapatilhas femininas ao comprar tênis e sandálias pode alavancar o ticket médio.
* **Criação de Kits Personalizados**: Combinar itens recorrentes em kits enxoval para recém-nascido.
* **Melhoria no Layout da Loja**: Em lojas físicas, posicionar os produtos associados próximos uns dos outros pode estimular vendas cruzadas.
* **Recomendações Automáticas em e-commerce**: Usar essas regras como base para sugestões automáticas na finalização da compra.