

PAULO RAFAEL DE MOURA PAES

**OTIMIZAÇÃO DE SORTIDO PARA RETALHISTA ALIMENTAR DE
PROXIMIDADE UTILIZANDO DATA MINING - UMA ANÁLISE EMPÍRICA**



UNIVERSIDADE DO ALGARVE

FACULDADE DE ECONOMIA

2022

PAULO RAFAEL DE MOURA PAES

**OTIMIZAÇÃO DE SORTIDO PARA RETALHISTA ALIMENTAR DE
PROXIMIDADE UTILIZANDO DATA MINING – UMA ANÁLISE EMPÍRICA**

Mestrado em Gestão de Marketing

Trabalho efetuado sob a orientação de:

PROFESSORA DOUTORA CÉLIA MARIA QUITÉRIO RAMOS

PROFESSORA DOUTORA DORA LÚCIA MIGUEL AGAPITO



UNIVERSIDADE DO ALGARVE

FACULDADE DE ECONOMIA

2022

**OTIMIZAÇÃO DE SORTIDO PARA RETALHISTA ALIMENTAR DE
PROXIMIDADE UTILIZANDO DATA MINING – UMA ANÁLISE EMPÍRICA**

Declaração de Autoria de Trabalho

Declaro ser o autor deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Paulo Rafael de Moura Paes

.....

(assinatura)

Direitos de cópia ou Copyright

© **Copyright: Paulo Rafael de Moura Paes**

A Universidade do Algarve tem o direito, perpétuo e sem limites geográficos, de arquivar e publicitar este trabalho através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, de o divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

AGRADECIMENTOS

Aos meus pais, Maria e Paulo, agradeço por me ensinarem a ser quem sou, por me apoiarem nas minhas escolhas e por se fazerem presentes mesmo na distância. É uma sorte imensa ter nascido um filho dessa família que criaram com tanto amor, carinho e atenção.

Ao meu irmão, Henrique, agradeço por ter sido o suporte na minha ausência, por ter sido uma inspiração para o meu crescimento e por me mostrar que na vida podemos ser melhores a cada dia.

À minha companheira, Carina, sou eternamente grato por ter sido incansável até aqui, pelas madrugadas de apoio nos estudos e por partilhar a vida comigo, sendo a minha parceira para todos os momentos.

Aos meus sogros, agradeço por serem como uma segunda família para mim, por todo o apoio em todos momentos difíceis e por me fazerem sentir em casa.

Aos meus colegas de curso e aos amigos que fiz em Portugal, por percorrerem comigo esta caminhada repleta de novidades e obstáculos.

Por fim, o meu muito obrigado às minhas orientadoras, professora doutora Célia Ramos e professora doutora Dora Agapito, por todo o suporte que me foi dado e pela atenciosamente maneira como o fizeram. Foi uma imensa felicidade tê-las escolhido para esta jornada.

RESUMO

O setor do retalho sofre frequentes mudanças, impulsionadas pelas necessidades dos consumidores. Um aumento significativo do número de produtos existentes no mercado impacta o processo de vendas. Especialmente, no retalho de proximidade, onde o consumidor deseja conveniência e agilidade, faz-se essencial otimizar os espaços disponíveis para a exposição de produtos.

A gestão de sortido surge neste contexto como um processo de seleção e organização dos produtos, de forma a atender a necessidade do consumidor. Portanto, a gestão do produto é, não só uma ação concorrencial, mas sim estratégica pois, ao definir um sortido, atende-se a algumas parcelas de clientes em detrimento de outras.

Esta dissertação tem o objetivo de analisar e identificar quais são os principais produtos a serem considerados no sortido de uma rede de retalho no formato de proximidade em Portugal, com lojas em Lisboa e Porto. Para tal, através do uso do software Orange, foram aplicados métodos de *data mining*, nomeadamente árvore de decisão (AD) e regras de associação (RA). Para além disso, o estudo exhibe uma revisão de literatura que discute os principais conceitos que permeiam e compõem a gestão de sortido.

A análise dos dados deste estudo permitiu identificar os atributos mais relevantes nas vendas registadas em base de dados pelo retalhista e discutir as relações mais importantes entre produtos frequentemente presentes em conjunto com outros nas vendas. Também são apresentadas estatísticas descritivas que possuem relevância nas decisões de negócio, podendo revelar-se importante para outras corporações do mesmo setor.

A discussão dos resultados relaciona os aspetos relevantes nas vendas do retalhista com as necessidades de negócio, no âmbito do mercado em que se enquadra. Os resultados demonstram como auxiliar a gestão de marketing na toma de decisão, bem como as contribuições que tais resultados podem trazer para a ciência, futuros estudos e limitações.

Palavras-Chave: Gestão de Sortido, *Data Mining*, Retalho, Árvore de Decisão, Regras de Associação.

ABSTRACT

The retail sector experiences frequent changes driven by consumer demand. A significant increase in the number of existing products on the market impacts the sales process. In proximity retail, consumer wants convenience and agility, being crucial to optimize the spaces available for the exhibition of products.

Assortment management appears in this context as a process of product selection and organization to meet consumer needs. Therefore, product management is not only a competitive action but a strategic one because defining an assortment is crucial to designate a target.

This dissertation aims to analyze and identify the main products to consider in the assortment of a proximity retail format in Portugal, with stores located in Lisbon and Porto. To this end, was used the Orange software with the application of *data mining* methods, namely decision tree (AD) and association rules (RA). In addition, the study presents a literature review that discusses the main concepts around the assortment management topic.

The data analysis identifies the most relevant attributes in the sales database and discusses the most important relationships between products often present with others in sales. The study also presents descriptive statistics relevant to business decisions and may prove the importance of other corporations in the same sector.

The discussion of the results connects the relevant aspects of the retailer's sales combined with the business demands, according to the scope of the market. The results demonstrate how to contribute to the decision-making process in marketing management dimensions and the contributions to limitations, science and future studies.

Keywords: Assortment Management, *Data Mining*, Retail, Decision Tree, Association Rules.

ÍNDICE GERAL

AGRADECIMENTOS.....	iii
RESUMO.....	iv
ABSTRACT.....	v
ÍNDICE GERAL.....	vi
ÍNDICE DE FIGURAS.....	viii
ÍNDICE DE TABELAS.....	ix
LISTA DE ABREVIATURAS.....	x
1. INTRODUÇÃO.....	1
1.1. Apresentação do tema/problema de investigação.....	1
1.2. Objetivos.....	2
1.3. Estrutura da dissertação.....	2
2. REVISÃO DA LITERATURA.....	3
2.1. A gestão de categorias no retalho.....	3
2.2. O sortido e sua gestão.....	3
2.3. Análise e definição do sortido.....	4
2.4. A segmentação.....	5
2.5. O sortido no retalho de proximidade.....	7
2.6. Processo de Gestão de Categorias do Sortido.....	7
2.7. A análise de dados no contexto de retalho.....	9
2.8. Estudos empíricos sobre definição do sortido no retalho.....	9
3. METODOLOGIA.....	11
3.1. A utilização de data mining como método de análise de dados no contexto de otimização de sortido para o retalho.....	11
3.2. Local de estudo e acesso.....	12
3.3. Método de <i>data mining</i>	12
3.3.1. Recolha de dados.....	14
3.3.2. Dimensão da base de dados.....	16
3.3.3. Preparação e transformação de dados.....	17
3.3.4. Aplicação de data mining.....	19
3.3.4.1. Aplicação de árvore de decisão.....	21
3.3.4.2. Aplicação de regras de associação.....	22

4. RESULTADOS.....	23
4.1. Descrição e caracterização da base de dados	23
4.2. Análise descritiva das variáveis	26
4.2.1. Público.....	26
4.2.2. Armazém/loja	32
4.3. Análise empírica a partir dos métodos de <i>data mining</i> aplicados	45
4.3.1. Árvore de decisão do público por classe social	45
4.3.2. Regras de associação de produtos por loja e mês	47
5. DISCUSSÃO DOS RESULTADOS.....	52
5.1. Discussão dos resultados das estatísticas descritivas.....	52
5.2. Discussão dos resultados da árvore de decisão	53
5.3. Discussão das regras de associação de produtos por loja e mês ..	54
5.4. Implicações teóricas	55
5.5. Implicações para a gestão de marketing.....	55
6. CONCLUSÕES	56
6.1. Principais conclusões do estudo	56
6.2. Limitações	59
6.3. Sugestões para estudos futuros.....	59
7. REFERÊNCIAS.....	60

ÍNDICE DE FIGURAS

Figura 3.1 - Processo de Descoberta de Conhecimento em Bancos de Dados (KDD) ..	14
Figura 3.2 - Estrutura da hierarquia de dados para sortido.....	15
Figura 3.3 - Exemplificação da estrutura de produtos da categoria de soft drinks.....	15
Figura 3.4 - Estrutura de árvore de decisão	21
Figura 4.1 - Vendas líquidas e vendas em quantidade, por loja (%)	23
Figura 4.2 - Vendas líquidas e vendas em quantidade, por família de produtos (%)	24
Figura 4.3 - Vendas líquidas e vendas em quantidade, por público (%)	24
Figura 4.4 - Quantidade de produtos vendidos por categoria (n)	25
Figura 4.5 - Parâmetros aplicados à árvore de decisão.....	45
Figura 4.6 - Árvore de decisão do público por classe social	46
Figura 4.7 - Árvore de decisão do público baixa.....	47
Figura 4.8- Parâmetros aplicados às regras de associação	48
Figura 4.9 - Regras de associação de produtos por loja e mês com confiança de 100%	49
Figura 4.10 - Regras de associação de produtos por loja e mês com confiança de 99,3%	50

ÍNDICE DE TABELAS

Tabela 2.1 Dimensão ideal de um sortido para uma loja de acordo com Rousseau (2014)	5
Tabela 2.2 - Variáveis de segmentação de mercados consumidores de acordo com Kotler e Armstrong (2014)	6
Tabela 3.1 - Características das lojas do retalhista estudado	16
Tabela 3.2 - Variáveis da base estudada.....	17
Tabela 3.3 - Estrutura da base de dados para análise de cesta por armazéns	18
Tabela 3.4 - Estrutura da base de dados para análise de árvore de decisão por público	18
Tabela 3.5 - Métodos descritivos e preditivos por tipo de aprendizagem	20
Tabela 4.1 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão do público “Baixa” em Vendas Líquidas (€)	26
Tabela 4.2 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão do público “Média/Baixa” em Vendas Líquidas (€)	27
Tabela 4.3 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão do público “Média” em Vendas Líquidas.....	29
Tabela 4.4 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão do público “Média/Alta” em Vendas Líquidas.....	30
Tabela 4.5 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja “Buraca” em Vendas Líquidas (€).....	33
Tabela 4.6 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja “Camarinha” em Vendas Líquidas (€)	34
Tabela 4.7 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja “Entroncamento” em Vendas Líquidas (€).....	35
Tabela 4.8 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja “Porto de Mós” em Vendas Líquidas.....	37
Tabela 4.9 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja “Coimbra” em Vendas Líquidas (€)	38
Tabela 4.10 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja “Carregado” em Vendas Líquidas (€)	40
Tabela 4.11 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja “Massamá” em Vendas Líquidas (€).....	41
Tabela 4.12 - Cruzamento entre Família de Produto, Público e Loja em termos de Média, Máxima Mínima e Total, efetuada nas Vendas Líquidas (€)	44

LISTA DE ABREVIATURAS

KDD	Descoberta de Conhecimento em Bancos de Dados
AD	Árvore de Decisão
RA	Regras de Associação

1. INTRODUÇÃO

1.1. Apresentação do tema/problema de investigação

Nos tempos mais recentes, o retalho, bem como todas as suas dimensões, passaram por mudanças expressivas, especialmente nas necessidades dos consumidores que, a cada dia se tornam mais personalizadas e complexas. O reflexo de tal transformação, é, de facto, o aumento significativo de número de produtos existentes no mercado e mudanças no processo de venda e na forma com que estas são conduzidas pelos retalhistas (Hekimoğlua, Sevim, Aksezer e Durmuş, 2019).

Neogrid (2018) e Pratas e Brito (2019) defendem que o retalho tem a análise de sortido como um processo extremamente importante para se definir quais das categorias que serão disponibilizadas pelo negócio. É essencial saber como otimizar os espaços disponíveis para exposição e escolher o produto que de fato atender a necessidade de compra.

Os retalhistas de proximidade, por sua vez, encaram um desafio ainda maior, pois além de estes possuírem espaço para exposição de produto em loja muito limitado, o consumidor que se dirige a este tipo de loja tem uma missão de compra motivada por conveniência, agilidade na compra e confiança para encontrar o produto que necessita (Gahinet e Cliquet, 2018).

Pelo facto de que estes desafios são oportunidades já investigadas e documentadas por autores citados no presente estudo, o problema de investigação relaciona-se com perceber, de forma empírica, como otimizar o sortido do produto em lojas de proximidade. Utilizar-se-á dados reais e métodos de *data mining*, no contexto de uma rede de retalho no formato de proximidade em Portugal, com sete lojas distribuídas entre Lisboa e Porto. Tratar-se-á de uma análise empírica dos dados sobre produtos comercializados por estas, a fim de definir um sortido otimizado a um número ideal de produtos, com base na propensão de compra dos clientes da rede, na rentabilidade, preço de seus produtos e espaço linear das lojas.

Apesar de terem sido publicados alguns estudos no tópico (Miller, Smith, McIntyre e Achabal (2010); Kwak, Duvvuri e Russell (2015); Hekimoğlua et al. (2019); Sethuraman,

Gázquez-Abad e Martínez-López (2022)), pouco foi estudado sobre a utilização de *data mining* no âmbito da definição e otimização de sortido para lojas de retalho, em concreto, de proximidade. E foi com base nesta ausência de estudos que este projeto de dissertação foi proposto. As implicações de gestão de conhecimento a serem descobertas são cruciais para a formação estratégica do retalhista na definição de sortido. Para além disso, o estudo tem a função de oferecer maiores bases comprovadas ao sector de atividade, ajudando na construção de conhecimentos para a área do retalho e academia.

1.2. Objetivos

O objetivo geral do presente estudo é analisar e identificar quais são os principais produtos a serem considerados no sortido e em ações promocionais direcionadas de uma rede de retalho no formato de proximidade em Portugal, com lojas em Lisboa e no Porto, a partir da utilização de técnicas de *data mining*.

Os objetivos específicos para este estudo podem ser definidos por:

- Segmentar as lojas do retalhista, de acordo com a performance de vendas de seus produtos;
- Identificar quais são os principais produtos vendidos para cada tipo de público;
- Compreender quais são os principais produtos a serem considerados pelo retalhista para as lojas, tendo em conta a probabilidade de estes serem vendidos em conjunto, bem como com a performance das vendas.

1.3. Estrutura da dissertação

A presente dissertação segue uma estrutura de seis capítulos. O primeiro capítulo aborda o problema de pesquisa, a relevância e a contribuição esperada do estudo e os objetivos da pesquisa. O segundo capítulo é dedicado à revisão de literatura, seção que apresenta a definição de sortido em contexto do retalho de proximidade, bem como as definições de análise, definição e otimização deste. O terceiro capítulo descreve a metodologia e justifica o desenho do modelo e recolha de dados para o estudo. O quarto capítulo exhibe os resultados obtidos e a respetiva análise. O quinto capítulo apresenta as discussões acerca dos resultados obtidos. Por fim, o sexto capítulo apresenta as conclusões, recomendações, implicações e limitações do estudo, juntamente com sugestões para pesquisas futuras.

2. REVISÃO DA LITERATURA

2.1. A gestão de categorias no retalho

Dentre os diversos processos que constituem a estrutura de uma empresa retalhista está a gestão de categorias, caracterizada pelo processo de classificar e gerir as categorias de produto como uma unidade de negócio “autónoma” na organização e tem como base o objetivo de criar valor para o consumidor, fornecedores e para a própria organização retalhista. Isto significa que as insígnias retalhistas não estão estruturadas somente como conjuntos de produtos individuais comercializados sobre um único estabelecimento, mas agrupa um conjunto de artigos de forma coerente (Pratas e Brito, 2019).

2.2. O sortido e sua gestão

Nos últimos anos, o negócio de retalho, bem como todas as suas dimensões, sofreu grandes mudanças, principalmente impulsionadas pelas necessidades dos consumidores que passaram a ser cada vez mais personalizadas e complexas, pelo aumento significativo do número de produtos existentes no mercado e pelas mudanças no processo de venda e na forma em que estas são conduzidas pelos retalhistas (Hekimoğlua *et al.*, 2019).

Para que ocorra a seleção e organização dos produtos, de forma a atender essa complexa necessidade do consumidor, o processo de gestão do sortido tem um importante papel. O sortido é constituído por produtos que se agrupam em famílias ou subfamílias, representando grupos de produto para responder a uma finalidade global (Rousseau, 2014). A seleção de sortido além de uma ação concorrencial é também um ato fundamental, pois, ao defini-lo, atende a algumas parcelas de clientes em detrimento de outras. A composição do sortido de uma loja não é e nem deve ser permanente, a dimensão deste varia com o tempo e evolui com as movimentações do mercado (Rousseau, 2014).

Rousseau (2014) também defende que, quanto à previsão da composição de um sortido, pode-se definir de forma empírica ou por meio de planos de sortido rigorosos e que vão ao pormenor. O mesmo autor define um plano de sortido como “a descrição pormenorizada das gamas a propor à clientela durante um determinado período”

(Rosseau, 2014: 175). Os fatores a serem considerados nesta definição, segundo o autor, são:

1. **A duração** – deve ser anual, semestral ou trimestral. E, para determinadas categorias de produto em que a procura tenha alguma constância, o plano de sortido deve ser ajustado ou revisto por meio de retificações pontuais, consoante as necessidades de mudanças identificadas pela organização retalhista. E, por último, para os produtos sazonais, recomenda-se elaborar planos especiais ou temporários.

2. **O conteúdo** – devem ser considerados todos os produtos, família por família de produto, todos os produtos relacionados e os devidos preços. Aquelos produtos considerados pelo retalhista como básicos poderão necessitar de uma atenção individual, uma vez que estes têm maior relevância nas vendas globais.

2.3. Análise e definição do sortido

O aumento significativo do número de produtos disponíveis no mercado traz aos retalhistas o desafio de definir um sortido mais adequado para apresentar aos consumidores no momento da compra. No retalho, a análise de sortido é crucial para a tomada de decisão sobre as categorias a serem oferecidas pelas lojas. De facto, de acordo com a Neogrid (2018), “para o varejo, a análise de sortimento é extremamente importante na hora de definir quais categorias serão disponibilizadas pelo negócio. Ocupar uma gôndola com produtos que não serão aproveitados é não só um desperdício de espaço como também de uma oportunidade de oferecer itens que possam trazer mais retorno ao investimento realizado”. Gôndola entende-se por um termo utilizado no mercado brasileiro para designar uma estante, ou conjunto de prateleiras de uma loja, onde estão expostos os bens para consumo.

Segundo Rousseau (2014), a dimensão de um sortido, em termos genéricos, é composta por três dimensões:

1. **Largura** - verifica o número de famílias e categorias de produto comercializados pelo ponto de venda, que reflete o número de necessidades do consumidor atendidas;
2. **Profundidade** – quantifica as referências ou artigos comercializados pelo ponto de venda em cada família ou categoria de produtos.

3. **Coerência** – refere-se à homogeneidade de um sortido de produtos, isto é, à ligação entre as necessidades do consumidor e cada produto definido para as atender. É também certo dizer que um sortido é classificado como coerente quando todas as famílias ou categorias de produto possuem profundidades aproximadamente iguais.

Ainda segundo Rousseau (2014), uma loja tem a sua personalidade e depende de variadas combinações das seguintes dimensões contidas na tabela 2.1.

Tabela 2.1 Dimensão ideal de um sortido para uma loja de acordo com Rousseau (2014)

Dimensão do sortido	Tipologias de lojas retalhistas
Largo e profundo	Hipermercados e grandes armazéns
Largo e pouco profundo	Supermercados médios ou hipermercados pequenos
Estreito e profundo	Lojas especializadas ou grandes superfícies especializadas
Estreito e pouco profundo	Lojas de conveniência e <i>hard discounts</i>

Fonte: Adaptado de Rousseau (2014).

2.4. A segmentação

No contexto atual do mercado de consumo, caracterizado principalmente pelo aumento dos diferentes perfis de consumidores, hábitos de compra e interesses, a segmentação tornou-se um dos fatores-chave para o sucesso dos profissionais de marketing, pois este processo é fundamental para a correta correspondência entre as necessidades dos consumidores e a capacidade das empresas as atenderem, permitindo que a empresa reaja de forma rápida e assertiva, aprimorando assim a sua vantagem competitiva. Neste sentido, operar um processo de segmentação eficaz é indispensável, nomeadamente para o setor do retalho, uma vez que o agrupamento de clientes com características de consumo semelhantes, proporciona à empresa a oportunidade de perceber melhor o seu objetivo de mercado, de customizar a sua abordagem e regulamentar os preços, produtos e/ou promoções direcionadas (Doğan, Ayçin e Bulut, 2018).

Kotler e Armstrong (2014) definem segmentação de mercado como o ato de dividir um mercado em segmentos de menor dimensão de consumidores com necessidades, características ou comportamentos diferentes, uma vez que estes podem necessitar de estratégias ou *marketing-mix* específicos para o seu perfil.

Os consumidores de qualquer mercado diferem ao nível dos seus desejos, recursos, localização geográfica, atitudes e práticas de compra e, através da segmentação, as empresas podem dividir mercados grandes e heterogéneos em segmentos menores que podem ser alcançados de forma mais eficaz com produtos e serviços que vão ao encontro das suas necessidades únicas. Não existe uma única forma de criar uma segmentação. Os profissionais de marketing devem explorar diversas variáveis de segmentação, sozinhas e combinadas, para encontrar a melhor forma de visualizar a estrutura de um mercado (Kotler e Armstrong, 2014).

Segundo Kotler e Armstrong (2014), as variáveis que podem ser usadas na segmentação de mercados de consumidores podem ser observadas na tabela 2.2.

Tabela 2.2 - Variáveis de segmentação de mercados consumidores de acordo com Kotler e Armstrong (2014)

Variável de segmentação	Características
Segmentação geográfica	Trata da divisão de um mercado em diferentes unidades geográficas, como regiões, estados, municípios e bairros. Uma empresa pode decidir operar em uma ou algumas áreas geográficas ou em todas elas, com atenção às diferenças de geografia, necessidades e desejos.
Segmentação demográfica	Divide os mercados em segmentos baseando-se nas variáveis tais como idade, estágio do ciclo de vida, género, renda, ocupação, educação, religião, etnia e geração, por exemplo. Os fatores demográficos são as bases mais populares para segmentar grupos de clientes.
Segmentação psicográfica	Divide os compradores com base em classe social, estilo de vida ou características de personalidade. Partindo do princípio de que pessoas do mesmo grupo demográfico podem ter características psicográficas muito diferentes.
Segmentação comportamental	Separa os compradores pelos seus conhecimentos, atitudes, usos ou respostas em relação ao produto. Muitos profissionais de marketing acreditam que as variáveis de comportamento são o melhor ponto de partida para construir segmentos.

Fonte: Adaptado de Kotler e Armstrong (2014).

A segmentação de mercado ajuda a entender as principais oportunidades de segmento de mercado da empresa. A partir disto, a corporação pode avaliar os vários segmentos e decidir quantos e quais segmentos ela pode atender melhor (Kotler e Armstrong, 2014).

2.5. O sortido no retalho de proximidade

Para os retalhistas de proximidade, a dificuldade em definir uma estratégia ao nível do sortido é ainda maior, pois além destes possuírem espaço para exposição de produto em loja muito limitado, o consumidor que aqui se dirige tem uma missão de compra motivada por conveniência, agilidade na compra e confiança para encontrar o produto que necessita (Gahinet e Cliquet, 2018). Assim, é crucial e uma vantagem competitiva para este tipo de retalhista entender quais são os produtos mais comprados pelos seus clientes e otimizar o número de produtos de acordo com a performance de cada um deles (Pratas e Brito, 2019).

2.6. Processo de Gestão de Categorias do Sortido

A implementação de gestão de categorias depende de alguns fatores, tais como: o nível de serviço e preço, vendas e rentabilidade das categorias de uma empresa retalhista (Pratas e Brito, 2019). Ainda segundo Pratas e Brito (2019), há oito passos que uma boa gestão de categorias deve seguir:

1. **Definição da categoria** – determinam-se os produtos individuais que vão compor a categoria, a partir da agregação dos produtos individuais em unidades base, subcategorias e categorias, as quais devem corresponder ao que o consumidor da insígnia retalhista necessita e deseja.
2. **Definição do papel da categoria** – considera e atende à importância da categoria de produto para o retalhista, fundamentada na sua importância para o consumidor, rentabilidade, frequência de compra e importância para o posicionamento da insígnia. As categorias podem ser classificadas como categorias de rotina, sazonal ou conveniência, de acordo com o posicionamento do retalhista no mercado para com a comercialização da categoria.
3. **Avaliação e desempenho da categoria** – cria-se um plano para a gestão da categoria com o objetivo de analisar a performance da categoria de produtos, as forças e fraquezas (baseadas em fatores internos do retalhista) e nas oportunidades e ameaças (baseadas em fatores externos do mercado onde o retalhista está inserido).
4. **Desempenho da categoria** – tende a ser medido através de *scorecards*, que identificam os indicadores-chaves da categoria, quantificando as oportunidades por

meio de análises de um conjunto de resultados, comparados com os objetivos fixados no início.

5. **Estratégia da categoria** – definem-se diretrizes para alavancar a performance da categoria de produtos. Esta deve estar alinhada à estratégia geral da insígnia retalhista e dar direcionamentos para o sortido de produto, promoções de vendas, estratégias de preço, critérios de exposição e visual merchandising do produto, bem como atender a aspetos de logística, abastecimento e jornada de compra (experiência do cliente no ponto de venda).

6. **Táticas da categoria** – determinam-se os aspetos operacionais dos elementos de *marketing-mix* da categoria de produto, assim como um conjunto de indicadores de performance que devem ser medidos e monitorados. Nesta etapa, definem-se as ações e calendarização para atingir a estratégia proposta no que tange o sortido por ponto de venda, estratégias de preço por ponto de venda e segmentação de clientes, racional de implementação dos produtos nos pontos de venda, comunicação, promoções e demais intervenções de marketing a serem direcionadas aos consumidores.

7. **Implementação** – cada um dos elementos definidos na fase “táticas da categoria” é implementado a partir da execução de tarefas específicas. É uma etapa que deve ser efetuada em parceria com os fornecedores, com um plano de negócios que atenda ao papel, estratégia e táticas da categoria de produtos.

8. **Revisão da categoria** – por fim, analisam-se os resultados obtidos e são projetadas ações de melhoria. Este processo deve ser contínuo e baseado nos históricos das ações aplicadas ao longo do tempo, na análise dos dados obtidos e na comparação entre a performance das lojas de forma a encontrar melhores práticas.

Nota-se que, entre os fatores mencionados, a definição do sortido é citada em vários momentos e etapas como um dos aspetos principais deste processo, para que a gestão de categorias se faça possível numa organização retalhista. E, para tal, também são necessárias análises de indicadores relacionados com fatores como as vendas, distribuição, abastecimento e exposição de produto para uma melhor tomada de decisão (Pratas e Brito, 2019).

2.7. A análise de dados no contexto de retalho

Para planear a otimização do sortido e orientá-lo para aquilo que é a real procura dos clientes de uma loja, é necessário obter diversas informações sobre a performance dos produtos nos retalhistas, através de análises de dados mais complexas. De forma geral, são usados dados originados a partir dos registos de vendas (tais como, quantidades e valor dos produtos vendidos, preço), dos tipos de produto (tais como, gama, marca, categoria) e do perfil das lojas (ex., público-alvo, tamanho, espaço linear das prateleiras, stock) (Pratas e Brito, 2019). Todos esses dados à disposição do retalhista, considerando as dimensões de produto, de canal (lojas) e tempo (período), resultam numa base de dados com grandes quantidades de linhas. Com o atual crescimento do número de lojas, produtos e clientes, estas bases de dados tornaram-se ainda maiores e mais complexas de se analisar (Bradlow, Gangwar, Kopalle, e Voleti, 2017).

De modo a tornar possível analisar uma base de dados de grande dimensão, as técnicas de *data mining* (conceito de mineração de dados) permitem o uso de métodos para identificar relacionamentos entre os dados através da verificação da existência de correlação entre os atributos. Assim, faz-se possível identificar um modelo que descreve dependências significativas entre os dados, através de regras que permitem efetuar previsões futuras significativas e potencialmente úteis a partir de grandes quantidades de dados disponíveis (Tekin, Etlioglu, Koyuncuoğlu e Tekin, 2019). No caso do retalho, as técnicas de *data mining* são utilizadas para identificar a previsão de vendas, que ajuda a examinar padrões e a tomar decisões sobre estoques; e a prever, com base nas compras de um item no período analisado e conjugado com regras de associação, a probabilidade do cliente comprar um item complementar (Rygielski, Wang, e Yen, 2002).

2.8. Estudos empíricos sobre definição do sortido no retalho

Miller, Smith, McIntyre e Achabal (2010) definiram uma metodologia para escolher os sortidos de retalho ideais para produtos comprados com pouca frequência e mensurar a robustez destes sortidos em relação à mudança na preferência dos compradores. Concluiu-se que há pouca abordagem ou modelagem rigorosa para escolher sortido ideal para categorias como estas e assim foi desenvolvido um método para avaliar a sensibilidade do lucro esperado do retalhista às mudanças nas preferências do comprador.

Kwak, Duvvuri e Russell (2015), nos Estados Unidos, examinaram o impacto da variedade do sortido no processo de escolha do comprador em loja, mais especificamente, verificaram que os consumidores exigem menos variedade para itens de qualidade na categoria de iogurtes. Para tal foi utilizado o método de incidência de categoria de logística multivariada e, como resultado, o estudo mostra que a escolha de iogurtes é afetada pelas percepções de qualidade da marca e que a redução da variedade tem impacto direto na percepção de qualidade.

Hekimoğlua *et al.* (2019), desenvolveram um esquema de otimização do sortido que maximiza a receita de um único supermercado que opera como filial de uma rede na Turquia. Criou-se um modo de regressão *log-linear* que representa a relação entre o atendimento e a procura por meio de dados históricos do ponto de venda.

Sethuraman, Gázquez-Abad e Martínez-López (2022) entenderam o efeito do tamanho do sortido do retalhista nas percepções do consumidor, escolha, vendas e participação. O estudo contou com a criação de meta-análise do sortido de 1936 efeitos de tamanho obtidos de 95 trabalhos académicos compostos por 177 estudos publicados durante 1970 e 2021. O principal resultado obtido mostra que a elasticidade-benefício positiva indicou que, em média, um sortido maior é benéfico para os compradores/retalhistas e que tal elasticidade-benefício do tamanho do sortido vem sendo maior em tempos mais recentes nos EUA, principalmente para produtos de mercearia. Também se especificou que a variedade de produtos nos EUA é maior do que nos países asiáticos e africanos.

A revisão de literatura permite observar que existem alguns artigos, em diferentes países, que estudaram o sortido de um retalhista. Estas análises foram feitas em perspetivas distintas, sendo uma análise da elasticidade do sortido, identificação do impacto do comportamento do comprador na escolha do sortido ou otimização do sortido para aumentar a receita do retalhista.

Os estudos encontrados são, em sua maioria focados em uma categoria de produtos específica de bens de consumo ou bens duráveis. Identificou-se uma escassez de estudos relacionados com a otimização de sortido que sejam focados em retalhistas de proximidade e que abordem diversas categorias, ao invés de apenas uma.

3. METODOLOGIA

De forma geral, a metodologia pode ter por base três abordagens: quantitativa, qualitativa e mista. A presente dissertação utiliza uma abordagem de natureza exploratória e descritiva, uma vez que se pretende analisar e identificar quais são os principais produtos a serem considerados no sortido de uma rede de retalho no formato de proximidade em Portugal.

A abordagem quantitativa utiliza números para avaliar um fenômeno ou ocorrência, com o objetivo de quantificar respostas ou registos em base de dados e interpretá-los para tomar decisões (Arghode, 2012). Foram procuradas e observadas as relações entre variáveis de modo a encontrar padrões e comportamentos que justifiquem e comprovem as respostas aos objetivos inicialmente propostos. E, portanto, e para a obtenção de conclusões mais concretas e próximas da realidade, foi utilizada a abordagem quantitativa.

3.1. A utilização de data mining como método de análise de dados no contexto de otimização de sortido para o retalho

Observou-se que há algumas publicações de estudos nos quais foram utilizados o *data mining* para o retalho como método de análise, com o objetivo de analisar o efeito nas vendas, tanto das ações promocionais como para entender o impacto da exposição de produto em loja.

Chen, Chen e Ching-Wen (2006) usaram técnicas de data mining para descobrir a relação implícita, mas significativa, entre a distância espacial relativa dos produtos expostos e as vendas unitárias dos itens em uma loja de retalho. O objetivo do esquema de mineração desenvolvido foi identificar e classificar os efeitos de tais relações. Neste artigo, o autor propôs um novo esquema de representação e desenvolveu um algoritmo baseado em análise de associação para dados empíricos de espaço, nas prateleiras do retalhista.

Moodley, Chiclana, Caraffini e Carterb (2020) desenvolveram um algoritmo de promoção direcionada para aumentar a eficácia das promoções pelos retalhistas do Reino Unido, encontrando conjuntos de itens frequentes. O autor propôs uma estrutura matemática para o direcionamento do conjunto de itens que é conduzido antes do direcionamento do cliente.

Hekimoğlu *et al.*, (2019) também utilizaram *data mining*, através de um modelo de programação não linear inteira mista, que visa maximizar o lucro em detrimento de espaço na prateleira. O método foi utilizado para otimização do sortido que maximiza a receita de um único supermercado que opera como filial de uma rede na Turquia.

Miller *et al.* (2010) utilizaram a programação linear inteira mista, também através de *data mining*, para escolher os sortidos de retalho ideais para produtos comprados com pouca frequência. Além disso, mensuraram a robustez destes sortidos em relação à mudança na preferência dos compradores, fazendo avaliação da categoria de aparelhos de DVD, para o retalho nos Estados Unidos.

3.2. Local de estudo e acesso

O estudo ocorreu em torno do contexto de negócio de uma rede de retalho portuguesa no formato de proximidade. Esta rede contém sete lojas distribuídas entre Lisboa e Porto. Para este presente estudo, o retalhista cedeu ao investigador uma base de dados de vendas para servir como objeto de estudo do problema de investigação em questão.

3.3. Método de *data mining*

De acordo com Sharma (2016), o *data mining* define-se por uma nova tecnologia com grandes poderes e potencial para apoiar as empresas a se concentrarem nas informações mais relevantes dos dados coletados sobre os seus clientes. Através dele é possível descobrir informações nos dados que os relatórios não podem revelar de maneira eficaz.

“*Data mining* (a etapa de análise do processo ‘Descoberta de Conhecimento em Bancos de Dados’, ou KDD), um subcampo interdisciplinar da ciência da computação, é o processo computacional de descobrir padrões em grandes conjuntos de dados envolvendo

métodos na interseção de inteligência artificial, aprendizado de máquina, estatísticas e sistemas de banco de dados. O objetivo geral do processo de mineração de dados é extrair informações de um conjunto de dados e transformá-las em uma estrutura compreensível para uso posterior” (Sharma e Sharma, 2016: 580). Ainda, segundo Sharma e Sharma (2016), o *data mining* tem como objetivo geral extrair informações de um conjunto de dados e transformá-las em uma estrutura compreensível para uso posterior e, compreender técnicas e algoritmos para determinar padrões interessantes de grandes conjuntos de dados.

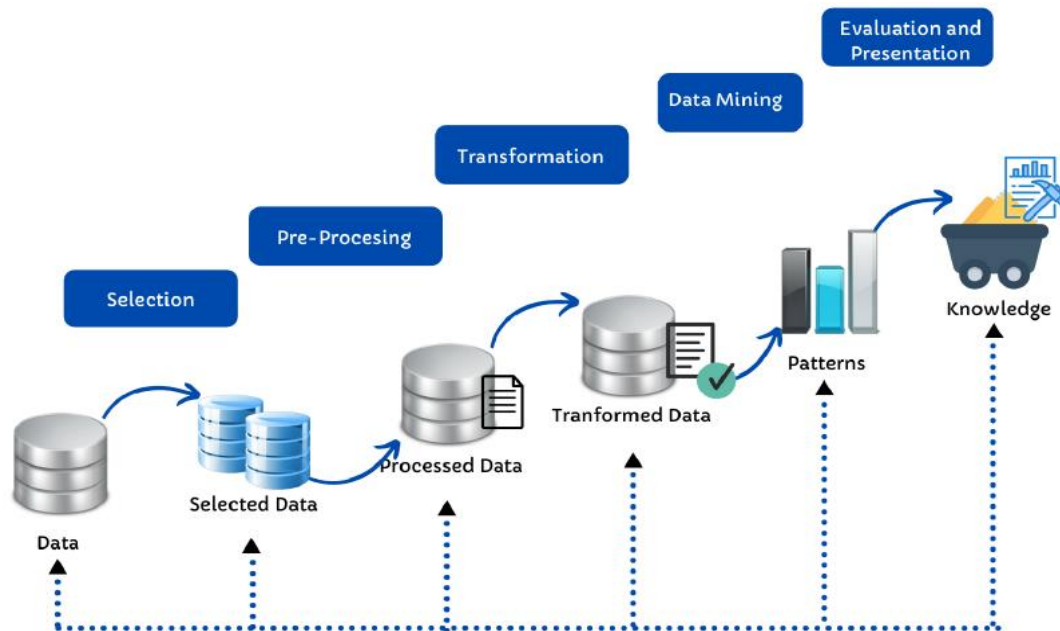
No retalho, a mineração de dados de vendas ajuda a identificar comportamentos de compra do cliente, melhorar a qualidade do atendimento ao cliente, alcançar melhor retenção e satisfação do cliente, melhorar as taxas de consumo de mercadorias, adequar as políticas de transporte e distribuição de mercadorias mais pungentes e reduzir o custo dos processos de negócios (Martinez, Escobar, Garcia-Diaz e Pinto-Roa, 2021).

Segundo Martinez *et al.* (2021) e Sharma e Sharma (2016), o processo de *data mining* é uma das etapas do processo de Descoberta de Conhecimento em Bancos de Dados (KDD), constituído por um conjunto de passos:

1. **Seleção de Dados:** os dados importantes para o pedido do analista são coletados no banco de dados.
2. **Pré-Processamento de Dados:** preparação dos dados, eliminando dados inconsistentes. É esta a fase mais exigente para que os dados sejam fiáveis.
3. **Transformação de dados:** os dados são transformados e consolidados de forma mais apropriada para mineração, através de resumos ou agregação dos dados.
4. **Data mining:** O processo essencial para extrair padrões de dados.
5. **Avaliação e Apresentação de Padrões:** etapa onde se identificam os padrões relevantes que representam o conhecimento com base em medidas que sejam interessantes. Para tal são utilizadas técnicas de visualização e representação do conhecimento, de forma que o usuário tenha entendimento do conhecimento extraído.

O conjunto de passos pode ser representado de forma gráfica pela Figura 3.1.

Figura 3.1 - Processo de Descoberta de Conhecimento em Bancos de Dados (KDD)



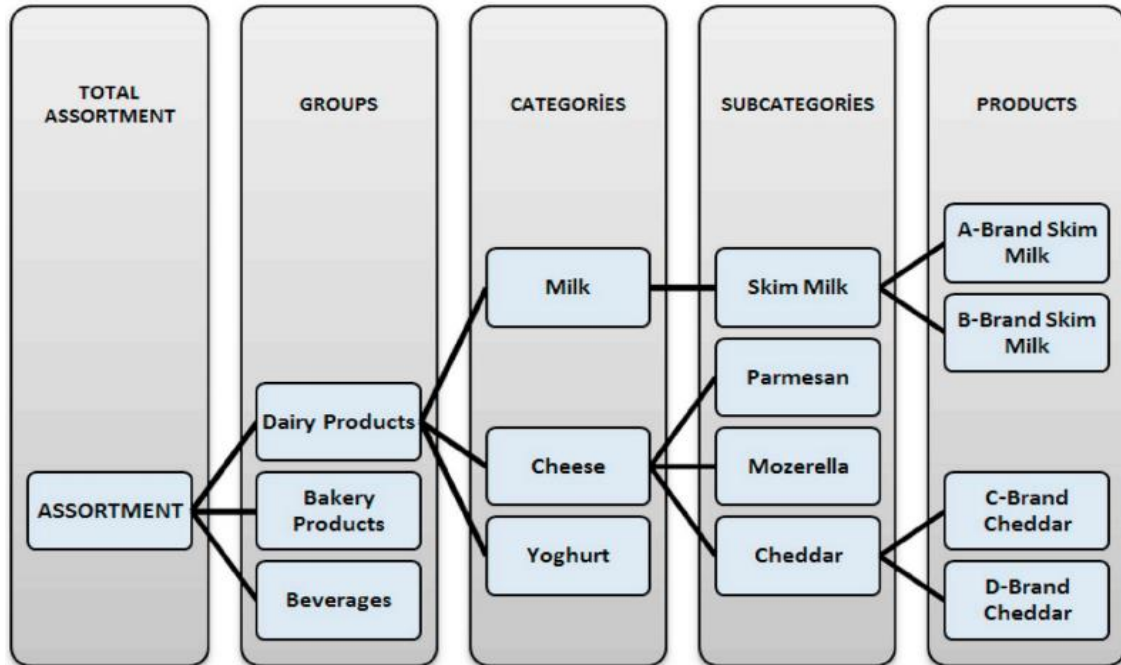
Fonte: Martinez *et al.* (2021: 4)

3.3.1. Recolha de dados

O investigador teve acesso à base de dados da empresa, incluindo os campos de vendas em quantidade e valor, vendas e preço para cada produto de todos os grupos de produtos, com uma periodicidade mensal durante 2 anos. Na base de dados, o sortido é dividido em diferentes grupos, como alimentos de mercearia, não comestíveis, frutas e vegetais, produtos domésticos e produtos de higiene pessoal. Esses grupos de produtos ramificam-se hierarquicamente em famílias, categorias, marcas e produtos. Exemplificando, os produtos lácteos, leite, leite desnatado e leite desnatado da Marca A podem ser classificados respetivamente em uma única categoria de lacticínios.

Os dados foram coletados a partir dos registos de vendas do retalhista no ponto de venda, que posteriormente foram disponibilizados em ficheiros no formato Microsoft Excel ao investigador.

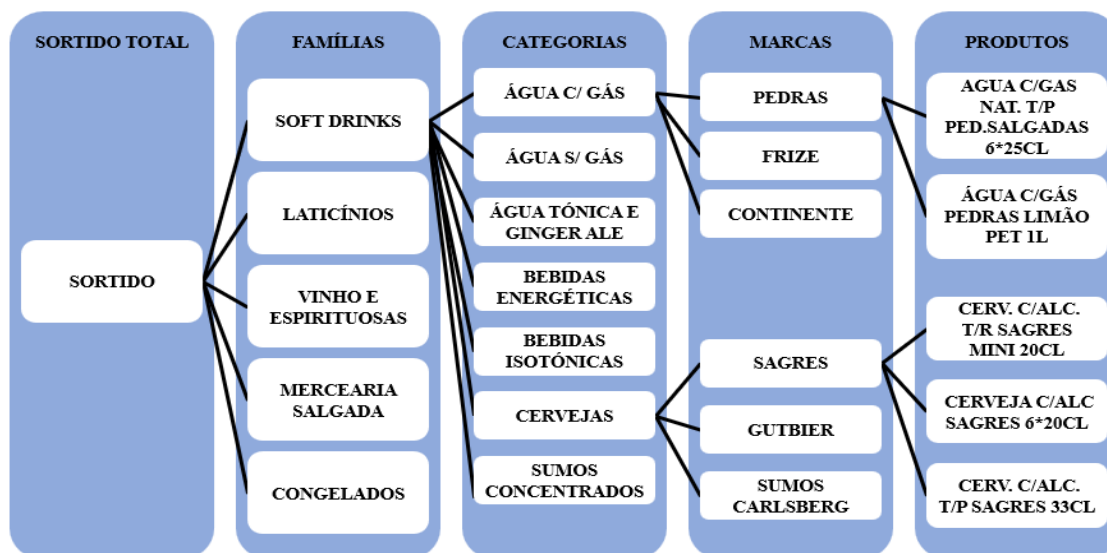
Figura 3.2 - Estrutura da hierarquia de dados para sortido



Fonte: Hekimoğlu *et al.* (2019: 202).

Uma exemplificação da estrutura de produtos da base de dados objeto de estudo pode ser visualizada na Figura 3.3, que resume a hierarquia da categoria de *soft drinks* e o objetivo é mostrar as suas ramificações. A mesma estrutura é aplicada nas demais categorias, que assumem os seus respetivos registos. Como referência foi utilizada a estrutura da hierarquia de dados para sortido, criada por Hekimoğlu *et al.* (2019: 202), exibida na Figura 3.2.

Figura 3.3 - Exemplificação da estrutura de produtos da categoria de soft drinks



Fonte: Adaptado de Hekimoğlu *et al.* (2019: 202).

3.3.2. Dimensão da base de dados

A base objeto de estudo contém dados de vendas dos produtos referentes às famílias de congelados, laticínios, mercearia salgada, *soft drinks*, vinhos e espirituosas. Esta possui 163.000 linhas de dados. As características das lojas consideradas no estudo estão descritas na Tabela 3.1

Tabela 3.1 - Características das lojas do retalhista estudado

Loja	Área	Público	Nº check-outs	Funcionários
Campo Maior	362	médio/alto	3	11
Camarinha	207	baixa	2	9
Buraca	105	baixa	2	5
Carregado	254	médio/baixo	2	8
Entroncamento	275	médio/baixo	2	8
Porto de Mós	240	médio	2	6
Coimbra	255	médio/alto	2	7

Fonte: Elaborado pelo autor.

As variáveis presentes na base objeto de estudo estão descritas na Tabela 3.2. Todas as variáveis são derivadas da base disponibilizadas e foram utilizadas na sua totalidade nas análises aplicadas no estudo.

Tabela 3.2 - Variáveis da base estudada

Variáveis	Referência	Tipo de variável
Armazém	Código do centro logístico que abastece a loja	Texto
Loja	Descrição da loja	Texto
Família	Código da família a qual o produto pertence	Numérica
Família (descrição)	Descrição do agrupamento de produtos ao qual o produto pertence	Texto
Artigo	Código do produto	Texto
Categoria	Descrição da categoria de produtos à qual o produto pertence	Texto
Marca	Descrição da marca à qual o produto pertence	Texto
Descrição	Descrição do produto	Texto
Quantidade	Quantidade vendida do produto	Numérica
V. Líquido	Valor (em euro) total da venda do produto	Numérica
Preço Médio	Preço médio (em euro) da venda do produto	Numérica
Mês	Mês em que o produto foi vendido	Data
Área de loja	Tamanho da loja em metros quadrados	Numérica
Público	Classe social predominante nos clientes da loja	Texto
Nº check-outs	Quantidade de caixas de atendimento da loja	Numérica
Funcionários	Número de funcionários da loja	Numérica

Fonte: Elaborado pelo autor.

3.3.3. Preparação e transformação de dados

De forma a tornar a base de dados objeto de estudo mais apropriada para o uso ao que foi proposto, foram necessários alguns ajustes, nomeadamente casas decimais que foram uniformizadas, limpeza e transformação, uma vez que foi utilizado o espaço como separador dos milhares, o qual teve de ser eliminada para que fosse possível efetuar as análises, bem como a exclusão de registos de vendas que se encontravam negativos ou zerados.

Neste sentido, também foram feitas transformações na estrutura da base específica para que fosse possível a análise de cesta por armazéns e por mês, onde foram concatenados os registos de armazém e mês de venda e colocada a informação de armazéns e meses de venda nas linhas (concatenados, de forma a criar um código único com a estrutura “loja_mês_ano”). Nas colunas foram colocados os produtos, sendo representada nos valores, a presença ou não de venda dos produtos nos respetivos armazém e mês

(identificada por 1 quando há presença e nulo quando não há. A Tabela 3.3 é uma extração de uma parte da base que exemplifica a estrutura de base utilizada na análise.

Tabela 3.3 - Estrutura da base de dados para análise de cesta por armazéns

Armazém_Mês	MOSTARDA SAVORA FRS 100G	NACHO QUEIJO CONTINENTE 150G	*ÁGUA S/GAS CONTINENTE 6*50CL	*ÁGUA S/GAS CONTINENTE 1.5L	*ÁGUA S/GAS CONTINENTE 33CL
Camarinha_1_2019	1				
Camarinha_10_2018	1				
Camarinha_12_2018	1				
Carregado_1_2019	1				
Carregado_11_2018	1				
Carregado_12_2018	1	1			
Carregado_2_2019	1				
Carregado_5_2019	1				
Carregado_6_2019	1				
Carregado_9_2018	1				
Coimbra_1_2019	1		1	1	1
Coimbra_10_2018	1	1	1	1	1
Coimbra_11_2018	1	1	1	1	1
Coimbra_12_2018	1	1	1	1	1
Coimbra_2_2019	1		1	1	1
Coimbra_3_2019	1		1	1	1

Fonte: Elaborado pelo autor.

Outras transformações foram feitas no âmbito da análise de árvore de decisão por tipo de público. Para tal, foram extraídas da base completa, com todos os dados, apenas os atributos público, marca, descrição, quantidade, preço médio, valor líquido e família. A Tabela 3.4 é uma extração de uma parte da base que exemplifica a estrutura utilizada na análise.

Tabela 3.4 - Estrutura da base de dados para análise de árvore de decisão por público

público	Marca	Descrição	Quantidade	Preço Médio	V. Líquido	Família (Descrição)
baixo	SAGRES	CERV. C/ALC. T/R SAGRES MINI 20CL	6132	0,34	2094,10	Soft Drinks
médio/baixo	CONTINENTE	OVOS CLASSE M CONTINENTE 1/2DZ	1011	0,84	849,80	Laticínios
médio/baixo	COCA COLA	REF.C/GAS COCA COLA 1,5L	519	1,37	711,25	Soft Drinks
baixo	PARMALAT	LEITE UHT M/G PARMALAT 1L	1546	0,45	700,38	Laticínios
médio/alto	CONTINENTE	OVOS CLASSE M CONTINENTE 1/2DZ	801	0,84	672,83	Laticínios
médio/baixo	CONTINENTE	+ÁGUA S/GAS CONTINENTE 6L	1170	0,56	650,35	Soft Drinks

médio/baixo	GUTBIER	CERV. C/ALC. TP GUTBIER 1L	833	0,76	630,78	Soft Drinks
médio/baixo	PRADO VERDE	LEITE UHT M/G PRADO VERDE 1L	1315	0,48	627,89	Laticínios
baixo	CONTINENTE	OVOS CLASSE M CONTINENTE 1/2DZ	726	0,84	609,84	Laticínios
baixo	SAGRES	CERV.C /ALC. T/R SAGRES 33CL	1176	0,50	583,11	Soft Drinks
baixo	SAGRES	CERV. C/ALC. T/P SAGRES 1L	377	1,54	579,79	Soft Drinks
baixo	MIMOSA	LEITE UHT M/G MIMOSA BEM ESSENCIAL 1LT	954	0,60	568,04	Laticínios

Fonte: Elaborado pelo autor.

3.3.4. Aplicação de data mining

Um estudo realizado na Turquia por Tekin *et. al* (2019) conclui que o *data mining* se tornou importante para a estratégia do marketing e para que os profissionais de marketing tomem decisões, a partir de previsões estratégicas para o futuro, tonando estas empresas mais competitivas. Segundo estes autores, muitas aplicações podem ser feitas especialmente na área de marketing, como determinar os hábitos de compra do cliente, determinar as relações entre as características demográficas dos clientes, aumentar a taxa de resposta da campanha, reter clientes existentes e adquirir novos clientes, aumentar a satisfação do cliente e reduzir o número de reclamações, facilitando o mercado análise de cesta, facilitando as práticas de gestão de relacionamento com o cliente, fazendo avaliações de clientes, ajudando a fazer previsões de vendas precisas e conduzindo várias análises de clientes.

É com base neste artigo e na análise dos demais pontos citados anteriormente, que se chegou a conclusão de que o método *data mining* é o mais adequado para o estudo a ser feito, pelo facto de que será necessário fazer previsões, relacionar e correlacionar variáveis e apresentar resultados para a tomada de decisão, fatores que são muito bem entregues pelo uso do método de análise, que permitirão a descoberta de padrões e/ou de relações entre os dados.

As atividades mais comuns em *data mining* podem ser divididas em dois tipos de métodos: o preditivo e o descritivo. Sendo o preditivo, as atividades realização por meio de inferências estatísticas nos dados de forma a fazer previsões relevantes, como por exemplo utilizar padrões de consumo para prever correlações entre os produtos e direcionar ações. Em relação às atividades descritivas, é feita uma classificação dos dados presentes na base a partir das suas propriedades, tornando-se possível descobrir padrões

e descrever os dados, para que os utilizadores possam fazer uma interpretação destes e tomar decisões (Martinez *et al.*, 2021).

Na Tabela 3.5 é possível ter uma melhor compreensão dos métodos descritivos e preditivos de *data mining* mais utilizados e suas principais características, divididos pelo tipo de aprendizagem ao qual pertencem.

Tabela 3.5 - Métodos descritivos e preditivos por tipo de aprendizagem

	Descritivo ou Preditivo	Método	Principais características
Aprendizagem supervisionada	Preditivo	Classificação e Regressão	Através da criação de modelos, este método é aplicado à custa de um conjunto de exemplos classificados, com o objetivo de prever para cada registo que faça parte de um outro conjunto de dados a pertença a uma determinada classe (classificação) ou um valor (regressão).
Aprendizagem não supervisionada	Descritivo	Clusterização (clustering)	Estas técnicas procuram semelhanças e diferenças em um conjunto de dados e agrupam estes registos semelhantes em segmentos ou clusters, automaticamente, utilizando de algum valor crítico ou métrica.
	Descritivo	Regras de associação	A abordagem em questão maneja a compra de um conjunto de produtos, assumindo como uma única transação. O objetivo é encontrar nas várias transações analisadas, tendências que possam ser consideradas para entender e explorar padrões de comportamento de compra.

Fonte: Adaptado de Oliveira (2000).

Para se fazerem operacionais as análises deste estudo, o software utilizado é o Orange (2022), software de *data mining* de livre acesso, que, com modelo pré-definidos de modelagem de dados e cálculos estatísticos, serve como base para as análises de *data mining* propostas nesta dissertação.

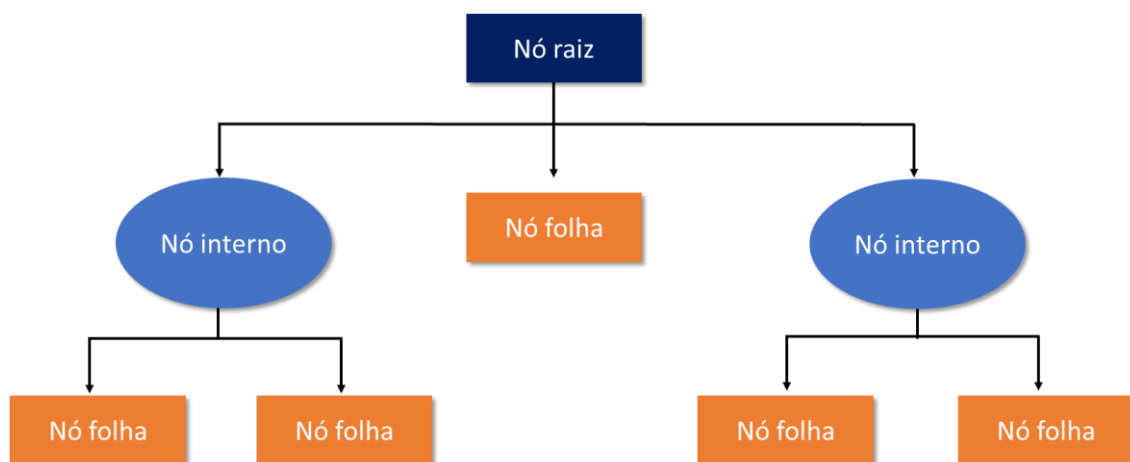
Segundo o site do fabricante, “Orange é a ferramenta perfeita para treinamento prático. Os professores apreciam o design claro do programa e as explorações visuais de dados e modelos. Os alunos se beneficiam da flexibilidade da ferramenta e do poder de inventar novas combinações de métodos de mineração de dados. A força educacional da Orange vem da combinação de programação visual e visualizações interativas. Também projetamos alguns *widjets* educacionais que foram criados explicitamente para apoiar o ensino” (Orange, 2022). Por ser de usabilidade adequada ao contexto de aprendizado que este estudo necessita.

Tendo em consideração o objetivo, os dados disponíveis e a necessidade de identificar os produtos mais vendidos por público e quais os produtos vendidos por loja e mês, os métodos de *data mining* a aplicar são AD (Árvore de Decisão) e RA (Regras de Associação), preditivo e descritivo respetivamente.

3.3.4.1. Aplicação de árvore de decisão

As árvores de decisão definem-se por métodos de aprendizado indutivos com uma estrutura de árvore hierárquica, conforme mostrado na Figura 3.4. Neste sentido, cada árvore pode ser de três tipos de nós: raiz, interno e folha. O nó raiz fica no topo de uma árvore e os nós internos ficam entre os nós raiz e folha. Cada um dos nós internos representa um teste em um atributo, cada ramificação do nó raiz até o nó folha significa a saída do teste e cada nó folha representa uma classificação de amostra (Jin, Li, Ma e Wang, 2022).

Figura 3.4 - Estrutura de árvore de decisão



Fonte: Adaptado de Jin *et al.* (2022: 4)

As árvores de decisão compõem uma classe de modelos preditivos com boas referências em eficiência e compreensão, especialmente devido à sua simplicidade. Estes modelos usam um processo em que os dados são decompostos em subconjuntos menores e mais puros, aplicando sistematicamente a pesquisa no espaço de possíveis ramificações das árvores de decisão e selecionando divisões ótimas com base em um critério de divisão. Esse processo começa no nó raiz, que é um nó sem nós raiz e determina iterativamente os critérios de divisão de ideais que dividem os dados em dois nós folha. O processo termina, quando outras divisões são desejáveis ou possíveis, com um conjunto de nós sem nós raiz chamados nós terminais ou folhas (Caigny, Coussement e Bock, 2018).

Uma vez que, substancialmente, um dos objetivos desta dissertação é encontrar os produtos mais vendidos por público, é utilizada a árvore de decisão, pois é possível identificar os nós derivados da venda por público dos produtos presentes na base de dados objeto de estudos. Desta forma, é possível entender quais categorias que são maioritariamente vendidas para cada nível de renda do público.

3.3.4.2. Aplicação de regras de associação

A mineração de regras de associação evidencia um trabalho vital em vários sistemas de tomada de decisão e ajudam a decidir as conexões entre os itens no banco de dados. Nesta metodologia, todos os itens assumem importância semelhante considerando a presença de itens em uma transação, não sendo considerada a utilidade dos itens. A mineração de regras de associação é de grande utilidade e um tema de estudo fascinante, que ajuda inúmeras empresas a analisar e diagnosticar problemas em seus negócios, suas tendências de vendas, bem como descobrir padrões de clientes, tomar possíveis decisões de negócios sobre a venda e gerar mais lucros (Krishna e Ravi, 2020).

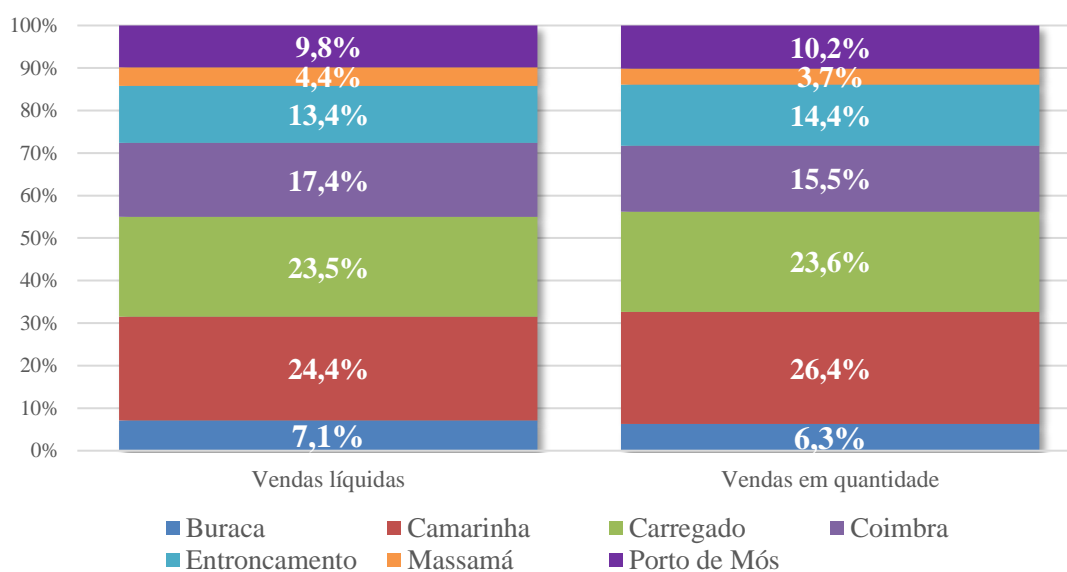
O entendimento de que conjuntos de produtos é vendido maioritariamente por loja e mês e quais produtos são provavelmente vendidos no mesmo período é uma das análises pretendidas nesta dissertação. Neste sentido, considera-se que as características do método de regra de associações definidas e citadas pelo autor citado no parágrafo anterior vão ao encontro do que se é desejado e dá um contributo para se obter os resultados.

4. RESULTADOS

4.1. Descrição e caracterização da base de dados

Foram obtidos dois anos de registos de vendas da empresa retalhista, recolhidos das sete lojas distribuídas por regiões de Portugal, situadas nas zonas de Buraca, Camarinha, Carregado, Entroncamento, Massamá e Porto de Mós. A concentração de tais vendas bem como as características destas foram consideradas para análises em diferentes perspetivas as quais podem ser analisadas nas figuras seguintes.

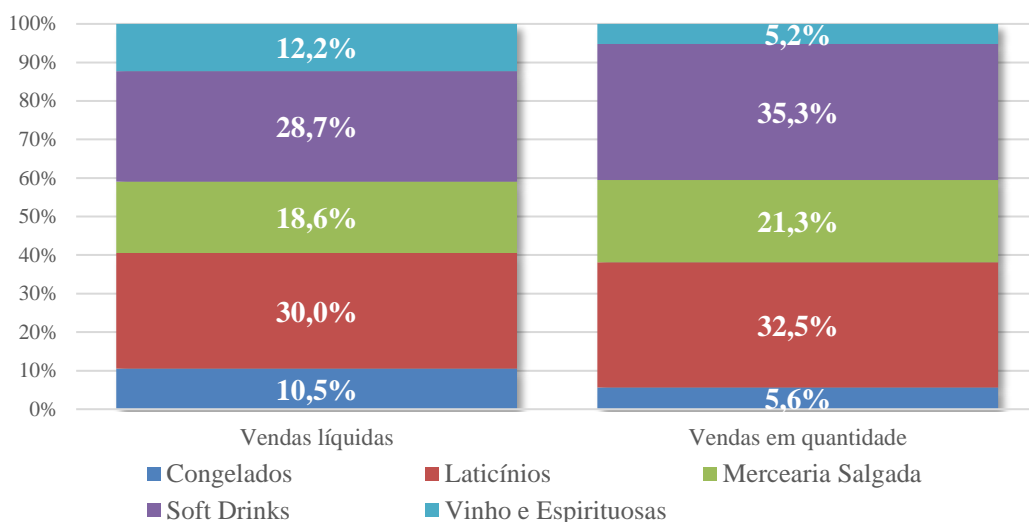
Figura 4.1 - Vendas líquidas e vendas em quantidade, por loja (%)



Fonte: Elaborado pelo autor.

É possível observar na Figura 4.1 que tanto as vendas líquidas quanto as vendas em quantidade do retalhista, provêm maioritariamente da loja Camarinha (24,4% e 26,4%, respetivamente), seguidos por Carregado (23,5% e 23,6%, respetivamente), que ocupa o segundo lugar e Coimbra (17,4% e 15,5%, respetivamente), em terceiro lugar.

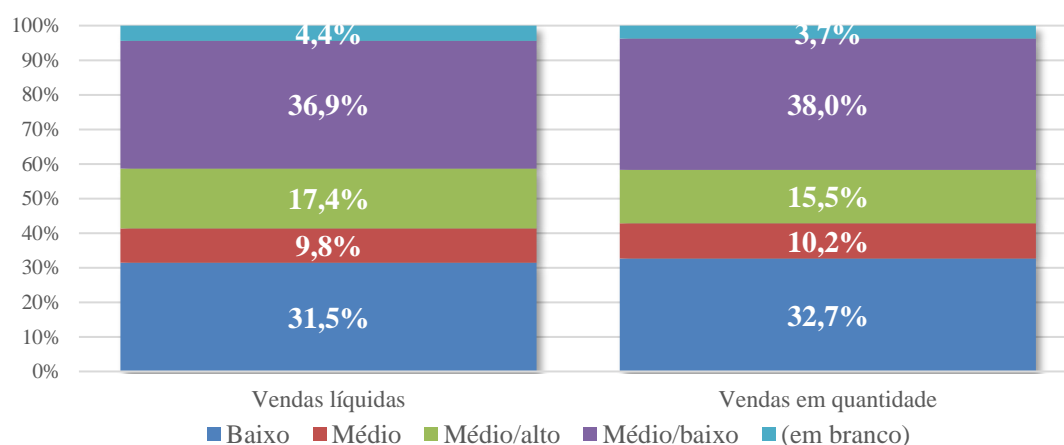
Figura 4.2 - Vendas líquidas e vendas em quantidade, por família de produtos (%)



Fonte: Elaborado pelo autor.

Na Figura 4.2, relativamente às vendas líquidas, a família de produtos líder é Laticínios (30,0%), seguida por *Soft Drinks* (28,7%) e Mercearia Salgada (18,6%). Em termos de vendas em quantidade, a família que lidera é *Soft Drinks* (35,3%), a segunda é Laticínios (32,5%) e mantém-se em terceiro a Mercearia Salgada (21,3%).

Figura 4.3 - Vendas líquidas e vendas em quantidade, por público (%)

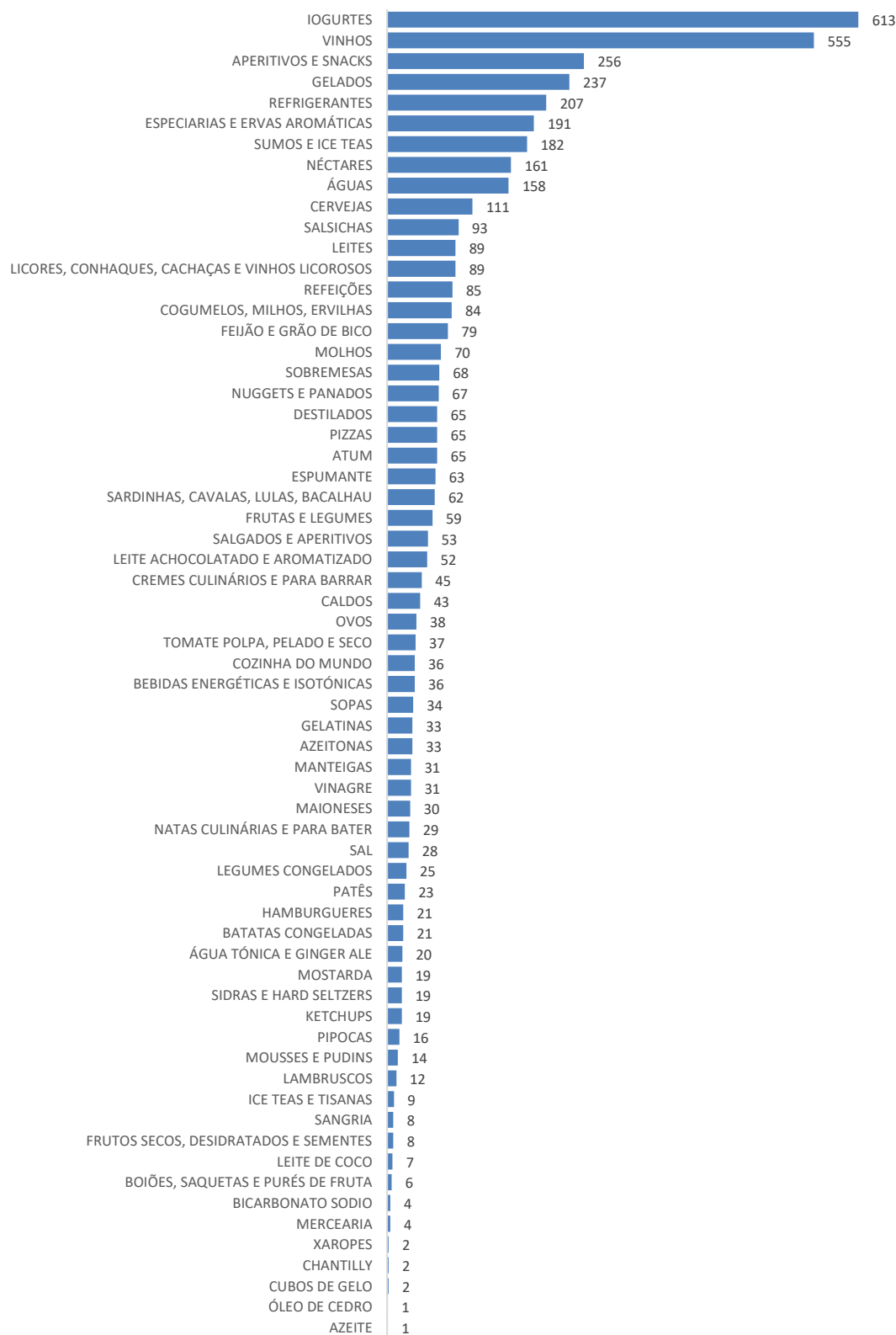


Fonte: Elaborado pelo autor.

Conforme os dados da Figura 4.3, o público que mais registou vendas na empresa retalhista no período analisado foi o público médio/baixo (36,9% em vendas líquidas e 38,0% em quantidade), seguido pelo público baixo (31,5% em vendas líquidas e 32,7% em quantidade) e por fim, o público médio/alto (17,4% em vendas líquidas e 15,5% em

quantidade), formando os três públicos mais representativos para a organização no aspeto de vendas.

Figura 4.4 - Quantidade de produtos vendidos por categoria (n)



Fonte: Elaborado pelo autor.

A quantidade de produtos vendidos por categoria à qual pertencem pode ser observada na Figura 4.4. O gráfico mostra que, as duas maiores categorias são Iogurtes, com um registo de vendas no período analisado de 613 produtos e Vinhos com um registo de vendas no período analisado de 555 produtos. A categoria de Azeites aparece como a menor representada em número de produtos, com somente 1 produto registado.

4.2. Análise descritiva das variáveis

4.2.1. Público

Observando a Tabela 4.1, relativamente à venda dos produtos pertencentes à família de congelados, a venda média para o público de baixa renda é de 22,18€. O valor mínimo vendido é 0,68€ e o valor máximo é de 346,93€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 3,50€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Tabela 4.1 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão do público "Baixa" em Vendas Líquidas (€)

	Congelados	Laticínios	Mercearia Salgada	Soft Drinks	Vinho e Espirituosas
Média	22,18	30,92	11,82	24,91	27,21
Moda	3,50	1,50	2,52	4,38	3,53
Máxima	346,93	877,18	744,24	2 094,10	1.128,35
Mínima	0,68	0,19	0,19	0,12	0,58
Desvio Padrão	31,23	66,69	20,75	81,05	53,62
Total	100.380,55	278 131,96	163.563,75	288.178,65	125.223,39

Fonte: Elaborado pelo autor.

Relativamente à venda dos produtos pertencentes à família de laticínios, a venda média de público de baixa renda é de 30,92€. O valor mínimo vendido é 0,19€ e o valor máximo é de 877,18€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 1,50€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de mercearia salgada, a venda média de público de baixa renda é de 11,82€. O valor mínimo vendido é 0,19€ e o valor máximo é de 744,24€, mostrando que as vendas do público se situam entre estes valores.

O valor que mais se repete é 2,52€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de *soft drinks*, a venda média de público de baixa renda é de 24,91€. O valor mínimo vendido é 0,12€ e o valor máximo é de 2.094,10€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 4,38€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de vinhos e espirituosas, a venda média de público de baixa renda é de 27,21€. O valor mínimo vendido é 0,58€ e o valor máximo é de 1.128,35€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 3,53€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

No que diz respeito à variabilidade das vendas observadas, as famílias de produto *soft drinks*, laticínios e vinhos e espirituosas são as que apresentam um maior desvio padrão, com valor de 81,05, 66,69 e 53,62, respetivamente. As famílias de produto congelados e mercearia salgada possuem um desvio padrão menor, sendo 31,23 e 20,75, respetivamente, o que representa menor variabilidade para estas famílias de produto.

Em termos de vendas líquidas totais, a família de produtos mais vendida para este público é a de *soft drinks* (288.178,65€), a segunda família mais vendida é a de laticínios (278.131,96€) e a terceira família mais vendida é a de mercearia salgada (163.563,75€). As duas famílias de produtos menos vendidas para este público são as de vinhos e espirituosas e congelados (125.223,39€ e 100.380,55€, respetivamente).

Tabela 4.2 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão do público "Média/Baixa" em Vendas Líquidas (€)

	Congelados	Laticínios	Mercearia Salgada	Soft Drinks	Vinho e Espirituosas
Média	19,46	29,52	12,67	23,79	21,03
Moda	3,37	2,64	1,68	5,07	2,48
Máxima	268,64	1.017,32	540,06	1.040,83	618,98
Mínima	0,48	0,14	0,16	0,14	0,29
Desvio Padrão	23,57	66,32	21,19	53,44	39,53
Total	111.457,88	313.345,45	206.264,90	300.627,45	99.667,41

Fonte: Elaborado pelo autor.

Observando a Tabela 4.2, relativamente à venda dos produtos pertencentes à família de congelados, a venda média de público de média/baixa renda é de 19,46€. O valor mínimo vendido é 0,48€ e o valor máximo é de 268,64€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 3,37€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de laticínios, a venda média de público de média/baixa renda é de 29,52€. O valor mínimo vendido é 0,14€ e o valor máximo é de 1.017,32€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 2,64€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de mercearia salgada, a venda média de público de média/baixa renda é de 12,67€. O valor mínimo vendido é 0,16€ e o valor máximo é de 540,06€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 1,68€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de *soft drinks*, a venda média de público de média/baixa renda é de 23,79€. O valor mínimo vendido é 0,14€ e o valor máximo é de 1.040,83€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 5,07€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de vinhos e espirituosas, a venda média de público de média/baixa renda é de 21,03€. O valor mínimo vendido é 0,29€ e o valor máximo é de 618,98€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 2,48€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

No que diz respeito à variabilidade das vendas observadas, as famílias de produto *soft drinks*, laticínios e vinhos e espirituosas são as que apresentam um maior desvio padrão, com valor de 53,44€, 66,32€ e 39,53€, respetivamente. As famílias de produto congelados

e mercearia salgada possuem um desvio padrão menor, sendo 23,57€ e 21,19€, respetivamente, o que representa menor variabilidade para estas famílias de produto.

Em termos de vendas líquidas totais, a família de produtos mais vendida para este público é a de laticínios (313.345,45€), a segunda família mais vendida é a de *soft drinks* (300.627,45€) e a terceira família mais vendida é a de mercearia salgada (206.264,90€). As duas famílias de produtos menos vendidas para este público são as de vinhos e espirituosas e congelados (99.667,41€ e 111.457,88€, respetivamente).

Tabela 4.3 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão do público "Média" em Vendas Líquidas

	Congelados	Laticínios	Mercearia Salgada	Soft Drinks	Vinho e Espirituosas
Média	10,61	19,82	8,33	12,77	12,83
Moda	4,22	7,80	1,26	1,46	1,76
Máxima	105,26	443,67	223,82	305,66	445,66
Mínima	0,52	0,19	0,21	0,13	0,29
Desvio Padrão	10,15	32,86	12,03	22,38	23,94
Total	28.105,25	99.815,98	62.253,53	75.731,19	31.255,32

Fonte: Elaborado pelo autor.

Observando a Tabela 4.3, relativamente à venda dos produtos pertencentes à família de congelados a venda média de público de média renda é de 10,61€. O valor mínimo vendido é 0,52€ e o valor máximo é de 105,26€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 4,22€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de laticínios a venda média de público de média renda é de 19,82€. O valor mínimo vendido é 0,19€ e o valor máximo é de 443,67€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 7,80€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de mercearia salgada a venda média de público de média renda é de 8,33€. O valor mínimo vendido é 0,21€ e o valor máximo é de 223,82€, mostrando que as vendas do público se situam entre estes valores.

O valor que mais se repete é 1,26€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de *soft drinks* a venda média de público de média renda é de 12,77€. O valor mínimo vendido é 0,13€ e o valor máximo é de 305,66€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 1,46€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.”

Relativamente à venda dos produtos pertencentes à família de vinhos e espirituosas a venda média de público de média renda é de 12,83€. O valor mínimo vendido é 0,29€ e o valor máximo é de 445,66€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 1,76€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

No que diz respeito à variabilidade das vendas observadas, as famílias de produto *soft drinks*, laticínios e vinhos e espirituosas são as que apresentam um maior desvio padrão, com valor de 22,38€, 32,86€ e 23,94€, respetivamente. As famílias de produto congelados e mercearia salgada possuem um desvio padrão menor, sendo 10,15€ e 12,03€, respetivamente, o que representa menor variabilidade para estas famílias de produto.

Em termos de vendas líquidas totais, a família de produtos mais vendida para este público é a de laticínios (99.815,98€), a segunda família mais vendida é a de *soft drinks* (75.731,19€) e a terceira família mais vendida é a de mercearia salgada (62.253,53€). As duas famílias de produtos menos vendidas para este público são as de vinhos e espirituosas e congelados (31.255,32€ e 28.105,25€, respetivamente).

Tabela 4.4 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão do público “Média/Alta” em Vendas Líquidas

	Congelados	Laticínios	Mercearia Salgada	Soft Drinks	Vinho e Espirituosas
Média	18,75	26,14	10,40	16,25	20,05
Moda	3,37	1,44	5,04	4,38	5,51
Máxima	165,78	672,83	330,00	683,50	550,87
Mínima	0,16	0,16	0,16	0,15	0,59
Desvio Padrão	17,13	45,94	15,50	31,25	30,83
Total	61.410,64	155.193,89	101.371,11	125.330,72	83.355,09

Fonte: Elaborado pelo autor.

Observando a Tabela 4.4, relativamente à venda dos produtos pertencentes à família de congelados a venda média de público de média renda é de 18,75€. O valor mínimo vendido é 0,16€ e o valor máximo é de 165,78€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 3,37€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de laticínios a venda média de público de média renda é de 26,14€. O valor mínimo vendido é 0,16€ e o valor máximo é de 672,83€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 1,44€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de mercearia salgada a venda média de público de média renda é de 10,40€. O valor mínimo vendido é 0,16€ e o valor máximo é de 330,00€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 5,04€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de *soft drinks* a venda média de público de média renda é de 16,25€. O valor mínimo vendido é 0,15€ e o valor máximo é de 683,50€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 4,38€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de vinhos e espirituosas a venda média de público de média renda é de 20,05€. O valor mínimo vendido é 0,59€ e o valor máximo é de 550,87€, mostrando que as vendas do público se situam entre estes valores. O valor que mais se repete é 5,51€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

No que diz respeito à variabilidade das vendas observadas, as famílias de produto *soft drinks*, laticínios e vinhos e espirituosas são as que apresentam um maior desvio padrão, com valor de 31,25€, 45,94€ e 30,83€, respetivamente. As famílias de produto congelados

e mercearia salgada possuem um desvio padrão menor, sendo 17,13€ e 15,50€, respetivamente, o que representa menor variabilidade para estas famílias de produto.

Em termos de vendas líquidas totais, a família de produtos mais vendida para este público é a de laticínios (155.193,89€), a segunda família mais vendida é a de *soft drinks* (125.330,72€) e a terceira família mais vendida é a de mercearia salgada (101.371,11€). As duas famílias de produtos menos vendidas para este público são as de vinhos e espirituosas e congelados (83.355,09€ e 61.410,64€, respetivamente).

Conclui-se que o público “Baixa” tem as maiores vendas líquidas médias entre os públicos, registando uma média mais expressiva na maioria das famílias de produtos. No entanto, o público “Média” é o público com menor expressão de vendas líquidas médias, sendo o público que regista as menores vendas líquidas médias em todas as famílias de produto.

Diferente das vendas médias, em termos de vendas totais, as conclusões são que o público que mais representa as vendas totais líquidas, em todas as famílias de produto, é a do público “Médio/baixo”. Neste sentido, o segundo público em vendas totais é o “Baixa” e o terceiro, o público “Médio/alto”. Ficando assim, o público “Médio”, como o quarto público em vendas totais.

4.2.2. Armazém/loja

Os dados presentes na Tabela 4.5, relativamente à venda dos produtos pertencentes à família de congelados dizem que a venda média da loja de Buraca é de 13,84€. O valor mínimo vendido é 0,70€ e o valor máximo é de 255,71€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 3,37€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de laticínios, os dados dizem que a venda média da loja de Buraca é de 19,42€. O valor mínimo vendido é 0,19€ e o valor máximo é de 477,56€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 0,33€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Tabela 4.5 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja “Buraca” em Vendas Líquidas (€)

	Congelados	Laticínios	Mercearia Salgada	Soft Drinks	Vinho e Espirituosas
Média	13,84	19,42	7,32	11,78	25,99
Moda	3,37	0,33	2,52	4,38	3,67
Máxima	255,71	477,56	186,00	335,12	371,40
Mínima	0,70	0,19	0,19	0,12	0,58
Desvio Padrão	21,58	37,36	11,78	18,39	44,12
Total	20.176,09	57.670,54	37.542,26	49.268,95	51.013,62

Fonte: Elaborado pelo autor.

Relativamente à venda dos produtos pertencentes à família de mercearia salgada, os dados dizem que a venda média da loja de Buraca é de 7,32€. O valor mínimo vendido é 0,19€ e o valor máximo é de 186,00€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 2,52€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de *soft drinks*, os dados dizem que a venda média da loja de Buraca é de 11,78€. O valor mínimo vendido é 0,12€ e o valor máximo é de 335,12€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 4,38€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de vinho e espirituosas, os dados dizem que a venda média da loja de Buraca é de 25,99€. O valor mínimo vendido é 0,58€ e o valor máximo é de 371,4€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 3,67€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

No que diz respeito à variabilidade das vendas observadas, as famílias de produto congelados, laticínios e vinhos e espirituosas são as que apresentam um maior desvio padrão, com valor de 21,58€, 37,36€ e 44,12€, respetivamente. As famílias de produto *soft drinks* e mercearia salgada possuem um desvio padrão menor, sendo 18,39€ e 11,78€, respetivamente, o que representa menor variabilidade para estas famílias de produto.

Em termos de vendas líquidas totais, a família de produtos mais vendida nesta loja é a de laticínios (57.670,54€), a segunda família mais vendida é a de vinhos e espirituosas (51.013,62€) e a terceira família mais vendida é a de *soft drinks* (49.268,95€). As duas famílias de produtos menos vendidas para este público são as de mercearia salgada e congelados (37.542,26€ e 20.176,09€, respetivamente).

Tabela 4.6 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja "Camarinha" em Vendas Líquidas (€)

	Congelados	Laticínios	Mercearia Salgada	Soft Drinks	Vinho e Espirituosas
Média	26,15	36,58	14,47	32,35	28,12
Moda	3,50	2,50	4,20	1,52	3,53
Máxima	346,93	877,18	744,24	2.094,10	1.128,35
Mínima	0,68	0,19	0,37	0,23	0,88
Desvio Padrão	34,19	76,52	24,15	99,73	59,71
Total	80.204,46	220.461,42	126.021,49	238.909,70	74.209,77

Fonte: Elaborado pelo autor.

Os dados presentes na Tabela 4.6, relativamente à venda dos produtos pertencentes à família de congelados dizem que a venda média da loja de Camarinha é de 26,15€. O valor mínimo vendido é 0,68€ e o valor máximo é de 346,93€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 3,50€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de laticínios, os dados dizem que a venda média da loja de Camarinha é de 36,58€. O valor mínimo vendido é 0,19€ e o valor máximo é de 877,18€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 2,50€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de mercearia salgada, os dados dizem que a venda média da loja de Camarinha é de 14,47€. O valor mínimo vendido é 0,37€ e o valor máximo é de 744,24€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 4,2€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de *soft drinks*, os dados dizem que a venda média da loja de Camarinha é de 32,35€. O valor mínimo vendido é 0,23€ e o valor máximo é de 2.094,10€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 1,52€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de vinhos e espirituosas, os dados dizem que a venda média da loja de Camarinha é de 28,12€. O valor mínimo vendido é 0,88€ e o valor máximo é de 1.128,35€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 3,53€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

No que diz respeito à variabilidade das vendas observadas, as famílias de produto *soft drinks*, laticínios e vinhos e espirituosas são as que apresentam um maior desvio padrão, com valor de 99,73€, 76,52€ e 59,71€, respetivamente. As famílias de produto congelados e mercearia salgada possuem um desvio padrão menor, sendo 34,19€ e 24,15€, respetivamente, o que representa menor variabilidade para estas famílias de produto.

Em termos de vendas líquidas totais, a família de produtos mais vendida nesta loja é a de *soft drinks* (238.909,70€), a segunda família mais vendida é a de laticínios (220.461,42€) e a terceira família mais vendida é a de mercearia salgada (126.021,49€). As duas famílias de produtos menos vendidas para este público são as de vinhos e espirituosas e congelados (74.209,77€ e 80.204,46€, respetivamente).

Tabela 4.7 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja "Entroncamento" em Vendas Líquidas (€)

	Congelados	Laticínios	Mercearia Salgada	Soft Drinks	Vinho e Espirituosas
Média	14,49	23,36	9,76	17,51	16,48
Moda	3,37	1,32	1,68	2,92	2,48
Máxima	119,42	601,60	334,31	547,64	476,61
Mínima	0,48	0,14	0,16	0,14	0,29
Desvio Padrão	14,11	46,04	14,27	32,21	33,76
Total	37.178,17	133.085,18	77.852,84	111.812,84	46.988,52

Fonte: Elaborado pelo autor.

Os dados presentes na Tabela 4.7, relativamente à venda dos produtos pertencentes à família de congelados dizem que a venda média da loja de Entroncamento é de 14,49€. O valor mínimo vendido é 0,48€ e o valor máximo é de 119,42€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 3,37€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de laticínios, os dados dizem que a venda média da loja de Entroncamento é de 23,36€. O valor mínimo vendido é 0,14€ e o valor máximo é de 601,60€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 1,32€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de mercearia salgada, os dados dizem que a venda média da loja de Entroncamento é de 9,76€. O valor mínimo vendido é 0,16€ e o valor máximo é de 334,31€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 1,68€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de *soft drinks*, os dados dizem que a venda média da loja de Entroncamento é de 17,51€. O valor mínimo vendido é 0,14€ e o valor máximo é de 547,64€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 2,92€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de vinhos e espirituosas, os dados dizem que a venda média da loja de Entroncamento é de 16,48€. O valor mínimo vendido é 0,29€ e o valor máximo é de 476,61€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 2,48€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

No que diz respeito à variabilidade das vendas observadas, as famílias de produto *soft drinks*, laticínios e vinhos e espirituosas são as que apresentam um maior desvio padrão, com valor de 32,21€, 46,04€ e 33,76€, respetivamente. As famílias de produto congelados e mercearia salgada possuem um desvio padrão menor, sendo 14,11€ e 14,27€, respetivamente, o que representa menor variabilidade para estas famílias de produto.

Em termos de vendas líquidas totais, a família de produtos mais vendida nesta loja é a de laticínios (133.085,18€), a segunda família mais vendida é a de *soft drinks* (111.812,84€) e a terceira família mais vendida é a de mercearia salgada (77.852,84€). As duas famílias de produtos menos vendidas para este público são as de vinhos e espirituosas e congelados (46.988,52€ e 37.178,17€, respetivamente).

Tabela 4.8 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja "Porto de Mós" em Vendas Líquidas

	Congelados	Laticínios	Mercearia Salgada	Soft Drinks	Vinho e Espirituosas
Média	10,61	19,82	8,33	12,77	12,83
Moda	4,22	7,80	1,26	1,46	1,76
Máxima	105,26	443,67	223,82	305,66	445,66
Mínima	0,52	0,19	0,21	0,13	0,29
Desvio Padrão	10,15	32,86	12,03	22,38	23,94
Total	28.105,25	99.815,98	62.253,53	75.731,19	31.255,32

Fonte: Elaborado pelo autor.

Os dados presentes na Tabela 4.8, relativamente à venda dos produtos pertencentes à família de congelados dizem que a venda média da loja de Porto de Mós é de 10,61€. O valor mínimo vendido é 0,52€ e o valor máximo é de 105,26€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 4,22€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de laticínios, os dados dizem que a venda média da loja de Porto de Mós é de 19,82€. O valor mínimo vendido é 0,19€ e o valor máximo é de 443,67€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 7,80€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de mercearia salgada, os dados dizem que a venda média da loja de Porto de Mós é de 8,33€. O valor mínimo vendido é 0,21€ e o valor máximo é de 223,82€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 1,26€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de *soft drinks*, os dados dizem que a venda média da loja de Porto de Mós é de 12,77€. O valor mínimo vendido é 0,13€ e o valor máximo é de 305,66€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 1,46€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de vinhos e espirituosas, os dados dizem que a venda média da loja de Porto de Mós é de 12,82€. O valor mínimo vendido é 0,29€ e o valor máximo é de 445,66€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 1,76€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

No que diz respeito à variabilidade das vendas observadas, as famílias de produto *soft drinks*, laticínios e vinhos e espirituosas são as que apresentam um maior desvio padrão, com valor de 22,38€, 32,86€ e 23,94€, respetivamente. As famílias de produto congelados e mercearia salgada possuem um desvio padrão menor, sendo 10,15€ e 12,03€, respetivamente, o que representa menor variabilidade para estas famílias de produto.

Em termos de vendas líquidas totais, a família de produtos mais vendida nesta loja é a de laticínios (99.815,98€), a segunda família mais vendida é a de *soft drinks* (75.731,19€) e a terceira família mais vendida é a de mercearia salgada (62.253,53€). As duas famílias de produtos menos vendidas para este público são as de vinhos e espirituosas e congelados (31.255,32€ e 28.105,25€, respetivamente).

Tabela 4.9 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja "Coimbra" em Vendas Líquidas (€)

	Congelados	Laticínios	Mercearia Salgada	Soft Drinks	Vinho e Espirituosas
Média	18,75	26,14	10,40	16,25	20,05
Moda	3,37	1,44	5,04	4,38	5,51
Máxima	165,78	672,83	330,00	683,50	550,87
Mínima	0,16	0,16	0,16	0,15	0,59
Desvio Padrão	17,13	45,94	15,50	31,25	30,83
Total	61.410,64	155.193,89	101.371,11	125.330,72	83.355,09

Fonte: Elaborado pelo autor.

Os dados presentes na Tabela 4.9, relativamente à venda dos produtos pertencentes à família de congelados dizem que a venda média da loja de Coimbra é de 18,75€. O valor mínimo vendido é 0,16€ e o valor máximo é de 165,78€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 3,37€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de laticínios, os dados dizem que a venda média da loja de Coimbra é de 26,14€. O valor mínimo vendido é 0,16€ e o valor máximo é de 672,83€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 1,44€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de mercearia salgada, os dados dizem que a venda média da loja de Coimbra é de 10,40€. O valor mínimo vendido é 0,16€ e o valor máximo é de 330€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 5,04€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de *soft drinks*, os dados dizem que a venda média da loja de Coimbra é de 16,25€. O valor mínimo vendido é 0,15€ e o valor máximo é de 683,50€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 4,38€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de vinho e espirituosas, os dados dizem que a venda média da loja de Coimbra é de 20,05€. O valor mínimo vendido é 0,59€ e o valor máximo é de 550,87€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 5,51€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

No que diz respeito à variabilidade das vendas observadas, as famílias de produto *soft drinks*, laticínios e vinhos e espirituosas são as que apresentam um maior desvio padrão, com valor de 31,25€, 45,94€ e 30,83€, respetivamente. As famílias de produto congelados e mercearia salgada possuem um desvio padrão menor, sendo 17,13€ e 15,50€, respetivamente, o que representa menor variabilidade para estas famílias de produto.

Em termos de vendas líquidas totais, a família de produtos mais vendida nesta loja é a de laticínios (155.193,89€), a segunda família mais vendida é a de *soft drinks* (125.330,72€) e a terceira família mais vendida é a de mercearia salgada (101.371,11€). As duas famílias de produtos menos vendidas para este público são as de vinhos e espirituosas e congelados (83.355,09€ e 61.410,64€, respetivamente).

Tabela 4.10 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja "Carregado" em Vendas Líquidas (€)

	Congelados	Laticínios	Mercearia Salgada	Soft Drinks	Vinho e Espirituosas
Média	23,50	35,54	15,47	30,53	24,63
Moda	3,37	1,68	1,68	5,07	4,59
Máxima	268,64	1.017,32	540,06	1.040,83	618,98
Mínima	0,68	0,21	0,39	0,17	0,88
Desvio Padrão	28,43	78,55	25,86	71,73	39,92
Total	74.279,71	201.045,27	128.444,09	237.918,28	69.051,52

Fonte: Elaborado pelo autor.

Os dados presentes na Tabela 4.10, relativamente à venda dos produtos pertencentes à família de congelados dizem que a venda média da loja de Carregado é de 23,50€. O valor mínimo vendido é 0,68€ e o valor máximo é de 268,64€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 3,37€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de laticínios, os dados dizem que a venda média da loja de Carregado é de 35,54€. O valor mínimo vendido é 0,21€ e o valor máximo é de 1.017,32€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 1,68€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de mercearia salgada, os dados dizem que a venda média da loja de Carregado é de 15,47€. O valor mínimo vendido é 0,39€ e o valor máximo é de 540,06€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 1,68€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de *soft drinks*, os dados dizem que a venda média da loja de Carregado é de 30,53€. O valor mínimo vendido é 0,17€ e o valor máximo é de 1.040,83€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 5,07€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de vinhos e espirituosas, os dados dizem que a venda média da loja de Carregado é de 24,63€. O valor mínimo vendido é 0,88€ e o valor máximo é de 618,98€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 4,59€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

No que diz respeito à variabilidade das vendas observadas, as famílias de produto *soft drinks*, laticínios e vinhos e espirituosas são as que apresentam um maior desvio padrão, com valor de 71,73€, 78,55€ e 39,92€, respetivamente. As famílias de produto congelados e mercearia salgada possuem um desvio padrão menor, sendo 28,43€ e 25,86€, respetivamente, o que representa menor variabilidade para estas famílias de produto.

Em termos de vendas líquidas totais, a família de produtos mais vendida nesta loja é a de *soft drinks* (237.918,28€), a segunda família mais vendida é a de laticínios (201.045,27€) e a terceira família mais vendida é a de mercearia salgada (128.444,09€). As duas famílias de produtos menos vendidas para este público são as de vinhos e espirituosas e congelados (69.051,52€ e 74.279,71€, respetivamente).

Tabela 4.11 - Média, Mediana, Moda, Máxima Mínima e Desvio Padrão da loja "Massamá" em Vendas Líquidas (€)

	Congelados	Laticínios	Mercearia Salgada	Soft Drinks	Vinho e Espirituosas
Média	12,12	18,16	6,86	10,68	11,67
Moda	3,50	3,68	1,31	1,47	8,77
Máxima	63,09	366,15	170,18	211,44	121,80
Mínima	0,71	0,19	0,22	0,17	0,61
Desvio Padrão	10,01	31,64	9,32	17,31	12,83
Total	17.796,54	39.930,96	29.104,51	31.429,90	14.544,57

Fonte: Elaborado pelo autor.

Os dados presentes na Tabela 4.11, relativamente à venda dos produtos pertencentes à família de Congelados dizem que a venda média da loja de Massamá é de 12,12€. O valor mínimo vendido é 0,71€ e o valor máximo é de 63,09€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 3,50€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de Laticínios, os dados dizem que a venda média da loja de Massamá é de 18,16€. O valor mínimo vendido é 0,19€ e o valor máximo é de 366,15€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 3,68€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de mercearia salgada, os dados dizem que a venda média da loja de Massamá é de 6,86€. O valor mínimo vendido é 0,22€ e o valor máximo é de 170,18€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 1,31€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de *soft drinks*, os dados dizem que a venda média da loja de Massamá é de 10,68€. O valor mínimo vendido é 0,17€ e o valor máximo é de 211,44€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 1,47€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

Relativamente à venda dos produtos pertencentes à família de vinhos e espirituosas, os dados dizem que a venda média da loja de Massamá é de 11,67€. O valor mínimo vendido é 0,61€ e o valor máximo é de 121,80€, mostrando que as vendas da loja se situam entre estes valores. O valor que mais se repete é 8,77€, ou seja, existe um valor de venda mais usual de produtos desse mesmo valor.

No que diz respeito à variabilidade das vendas observadas, as famílias de produto *soft drinks*, laticínios e vinhos e espirituosas são as que apresentam um maior desvio padrão, com valor de 17,31€, 31,64€ e 12,83€, respetivamente. As famílias de produto congelados e mercearia salgada possuem um desvio padrão menor, sendo 10,01€ e 9,32€, respetivamente, o que representa menor variabilidade para estas famílias de produto.

Em termos de vendas líquidas totais, a família de produtos mais vendida nesta loja é a de laticínios (39.930,96€), a segunda família mais vendida é a de *soft drinks* (31.429,90€) e a terceira família mais vendida é a de mercearia salgada (29.104,51€). As duas famílias de produtos menos vendidas para este público são as de vinhos e espíritos e congelados (14.544,57€ e 17.796,54€, respetivamente).

A análise descritiva apresentada mostra que, entre as lojas relacionadas, aquela que possui maiores vendas médias líquidas na maioria das famílias de produtos é a loja de Camarinha, com média superior às demais lojas nas famílias de produto congelados, laticínios, *soft drinks* e vinhos e espíritos. E, opostamente, a loja de Massamá mostra vendas médias líquidas com valores menores que as outras lojas na maior parte das famílias de produtos, sendo que este comportamento é verificado nas famílias de produto laticínios, mercearia salgada, *soft drinks* e, vinhos e espíritos.

Assim como na análise acima, que considera as vendas médias, em termos de vendas totais, as conclusões são que na maioria das famílias de produtos é a loja de Camarinha a loja com maiores vendas. A loja com menor representatividade em vendas é a loja de Massamá, que tem esse posto tanto na análise por vendas médias, quanto por vendas totais.

A tabela 4.12 permite uma avaliação global do retalhista a partir dos principais atributos de produto e loja. Neste sentido, é possível observar que, nas famílias de produtos de congelados, laticínios e *soft drinks*, a loja com melhor desempenho é a loja de Camarinha, uma vez que é a maior em termos de vendas totais. Na família de mercearia salgada, a maior venda total vem da loja Carregado e, na família de produtos de vinho e espíritos, a loja com maior venda total é a de Coimbra.

Ainda na tabela 4.12, em termos de vendas médias, as famílias de produto de congelados, laticínios, mercearia salgada e vinho e espíritos têm maiores valores vindos da loja de Camarinha. Apenas a família de produtos de *soft drinks* demonstra vendas médias maiores vindas da loja de Carregado.

Tabela 4.12 - Cruzamento entre Família de Produto, Público e Loja em termos de Média, Máxima Mínima e Total, efetuada nas Vendas Líquidas (€)

Famílias de produto	Loja	Público	Média	Máxima	Mínima	Total
Congelados	Buraca	baixa	13,84	255,71	0,70	20 176,09
Congelados	Camarinha	baixa	26,15	346,93	0,68	80 204,46
Congelados	Carregado	médio/baixo	23,50	268,64	0,68	74 279,71
Congelados	Coimbra	médio/alto	18,75	165,78	0,16	61 410,64
Congelados	Entroncamento	médio/baixo	14,49	119,42	0,48	37 178,17
Congelados	Massamá	médio/baixo	12,12	63,09	0,71	17 796,54
Congelados	Porto de Mós	médio	10,61	105,26	0,52	28 105,25
Laticínios	Buraca	baixa	19,42	477,56	0,19	57 670,54
Laticínios	Camarinha	baixa	36,58	877,18	0,19	220 461,42
Laticínios	Carregado	médio/baixo	35,54	1 017,32	0,21	201 045,27
Laticínios	Coimbra	médio/alto	26,14	672,83	0,16	155 193,89
Laticínios	Entroncamento	médio/baixo	23,36	601,60	0,14	133 085,18
Laticínios	Massamá	médio/baixo	18,16	366,15	0,19	39 930,96
Laticínios	Porto de Mós	médio	19,82	443,67	0,19	99 815,98
Mercearia Salgada	Buraca	baixa	7,32	186,00	0,19	37 542,26
Mercearia Salgada	Camarinha	baixa	14,47	744,24	0,37	126 021,49
Mercearia Salgada	Carregado	médio/baixo	15,47	540,06	0,39	128 444,09
Mercearia Salgada	Coimbra	médio/alto	10,40	330,00	0,16	101 371,11
Mercearia Salgada	Entroncamento	médio/baixo	9,76	334,31	0,16	77 852,84
Mercearia Salgada	Massamá	médio/baixo	6,86	170,18	0,22	29 104,51
Mercearia Salgada	Porto de Mós	médio	8,33	223,82	0,21	62 253,53
Soft Drinks	Buraca	baixa	11,78	335,12	0,12	49 268,95
Soft Drinks	Camarinha	baixa	32,35	2 094,10	0,23	238 909,70
Soft Drinks	Carregado	médio/baixo	30,53	1 040,83	0,17	237 918,28
Soft Drinks	Coimbra	médio/alto	16,25	683,50	0,15	125 330,72
Soft Drinks	Entroncamento	médio/baixo	17,51	547,64	0,14	111 812,84
Soft Drinks	Massamá	médio/baixo	10,68	211,44	0,17	31 429,90
Soft Drinks	Porto de Mós	médio	12,77	305,66	0,13	75 731,19
Vinho e Espirituosas	Buraca	baixa	25,99	371,40	0,58	51 013,62
Vinho e Espirituosas	Camarinha	baixa	28,12	1 128,35	0,88	74 209,77
Vinho e Espirituosas	Carregado	médio/baixo	24,63	618,98	0,88	69 051,52
Vinho e Espirituosas	Coimbra	médio/alto	20,05	550,87	0,59	83 355,09
Vinho e Espirituosas	Entroncamento	médio/baixo	16,48	476,61	0,29	46 988,52
Vinho e Espirituosas	Massamá	médio/baixo	11,67	121,80	0,61	14 544,57
Vinho e Espirituosas	Porto de Mós	médio	12,83	445,66	0,29	31 255,32

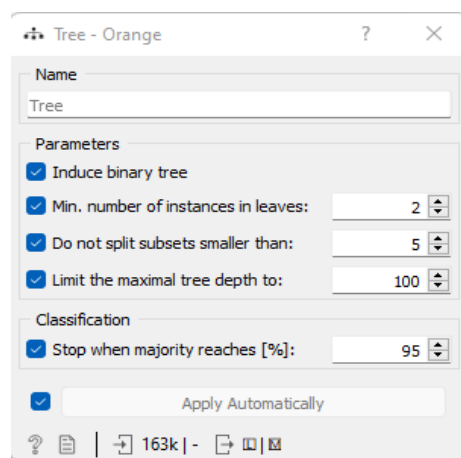
Fonte: Elaborado pelo autor.

4.3. Análise empírica a partir dos métodos de *data mining* aplicados

4.3.1. Árvore de decisão do público por classe social

A árvore de decisão contruída através da aplicação Orange baseada na estrutura de base já mencionada, leva em consideração os seguintes parâmetros que podem ser observados na Figura 4.5. Estes foram aplicados no *Widget Tree* da aplicação, onde são efetuados os cálculos e cruzamentos de dados necessário para esta análise.

Figura 4.5 - Parâmetros aplicados à árvore de decisão



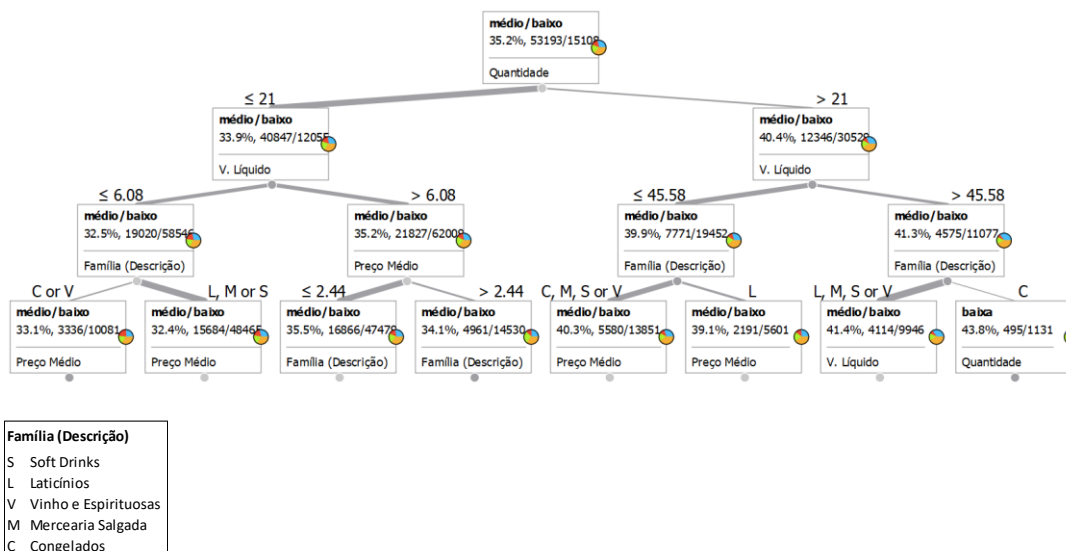
Fonte: Software Orange Data Mining

Ao selecionar a opção *Induce binary tree*, a aplicação constrói uma árvore binária (dividida em dois nós folha). O número mínimo de instâncias em folhas é igual a 2, o que faz com que o algoritmo nunca contrua uma divisão que colocaria menos do que 2 exemplos de treinamento em qualquer uma das ramificações. Também foi configurado não dividir subconjuntos menores que 5 nós. E, por último, limitou-se o algoritmo para que a profundidade da árvore de classificação seja de até 100.

A seleção de dados efetuada para esta análise, consiste em considerar os atributos Quantidade, Preço Médio, Valor Líquido e Família (Descrição) como características dos dados analisados. O atributo Público é considerado como variável target, pois é nele que está o objetivo desta análise, o alvo a ser considerado em todas as instâncias. Os atributos metas (nomenclatura definida pela aplicação para o atributo central da análise) desta análise são Marca e Descrição, pois são incluídos no conjunto de dados e no método, porém não são considerados na análise.

Após cumprir todas as etapas de *data mining*, o resultado da análise pode ser observado na Figura 4.6.

Figura 4.6 - Árvore de decisão do público por classe social



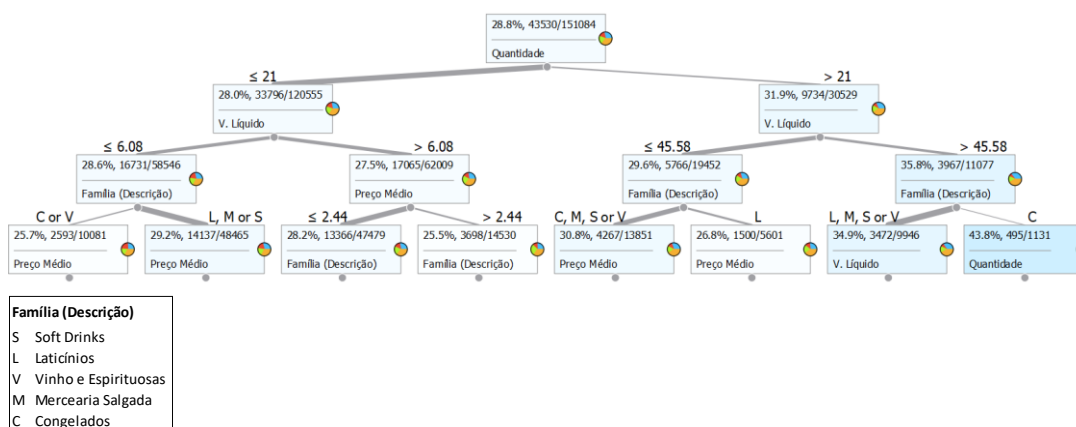
Fonte: Elaborado pelo autor.

Na árvore de decisão apresentada observa-se que o atributo Público é o objetivo da análise, pois é deste atributo que é desejado extrair informações que descrevam o seu comportamento na venda. Entre os resultados, é possível observar que, no primeiro nível, as vendas para o público médio/baixo representam (35,2%) e são a maior relação encontrada. Sendo que, para este público, quando são vendidos acima de 21 produtos (segundo nó) para o mesmo público, tais vendas representam 40,4% da Quantidade. No terceiro nó, é notado que as vendas para este mesmo público, em Valor Líquido, quando acima de 45,58 €, representam 41,3%. E, no quarto nível, é de notar que o público que cumpre este mesmo comportamento e a venda é majoritariamente para a família de produtos Congelados é o público baixa. Já quando são vendidas majoritariamente as famílias Laticínios, Mercearia Salgada, *Soft Drinks* ou Vinhos e Espirituosas, o público predominante é o médio/baixo.

Uma vez que o público observado no quarto nível da análise anterior sofre uma mudança de comportamento, saindo da predominância do público médio/baixo e mostrando resultados majoritários para o público baixo quando a família de produtos que predomina

é a de congelados, foi considerado relevante observar a árvore decisão quando a classe objetivo é o público baixa. Na Figura 4.7 é possível observar tal análise.

Figura 4.7 - Árvore de decisão do público baixa



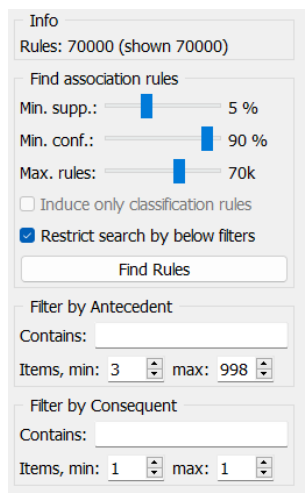
Fonte: Software Orange Data Mining

Na análise para o público baixa, no primeiro nível é possível observar que as vendas para este público representam 28,8%. No segundo nível, as vendas em Quantidade que representam acima de 21 produtos são 31,9% para este público e as vendas registradas com valor de Valor Líquido acima de 45,58 € representam 35,8% (terceiro nível). Em relação às famílias de produto para vendas com o mesmo comportamento observado nos outros níveis, as vendas registradas para a família Congelados é maioritária, sendo esta de 43,8% e, quando as vendas são para os produtos das famílias Laticínios, Merceria Salgada, *Soft Drinks* ou Vinhos e Espirituosas, a representação é de 34,9%.

4.3.2. Regras de associação de produtos por loja e mês

As regras de associação foram aplicadas, através da aplicação Orange, ao conjunto de vendas caracterizadas por loja e mês para todo o período de vendas presentes (identificados por 1) da base de dados estudada. Para tal, foram aplicados os parâmetros da Figura 4.8. Estes foram aplicados ao *wight Association Rules* da aplicação, onde foram configurados os parâmetros ao qual a análise deve respeitar.

Figura 4.8- Parâmetros aplicados às regras de associação



Fonte: Software Orange Data Mining

Aplicou-se o suporte mínimo, ou seja, a percentagem de todo o conjunto de dados coberto por toda a regra (antecedente e consequente), um valor mínimo de 5%. Neste sentido, permitiu que o mínimo de 90% de confiança na estatística aplicada e um máximo de 70.000 linhas de resultados a serem apresentados. Também foi configurado que a análise se restrinja a exibir apenas resultados que respeitem o filtro de 3 itens mínimos e 998 itens máximos em conjunto no antecedente e somente 1 item no consequente.

A seleção de dados criada para esta análise consiste em considerar como características todas as categorias de produto presentes na base de dados, representadas pela presença de venda na loja e mês referentes, e uma variável derivada da junção de loja e mês como meta, pois é incluída no conjunto de dados e no método, porém não são considerados na análise.

As teorias assumidas como embasamento para a criação da regra de associação, segundo a ferramenta Orange foram baseadas nos artigos de Han, Pei, Yin e Mao (2004) e Agrawal, Aggarwal e Prasad (2000).

Assim sendo, uma extração da análise em concreto pode ser visualizada na Figura 4.9, considerando associações com nível de confiança de 100% (ou 1.000, representado na figura), e na Figura 4.10, considerando um nível de confiança de 99,3% (ou 0.993, representado na figura).

A tabela à direita exibe uma lista de regras com 6 medidas diferentes de qualidade da regra de associação:

- **Supp:** trata-se de com que frequência uma regra é aplicável a um determinado conjunto de dados.
- **Conf:** representa com que frequência a regra é verdadeira (suporte para uma regra/suporte para antecedente).
- **Covr:** resulta da frequência em que o item antecedente é encontrado no conjunto de dados (suporte de antecedente/dados).
- **Strg:** (suporte de consequente/suporte de antecedente)
- **Lift:** exibe com que frequência uma regra faz-se verdadeira por item consequente (dados * confiança/suporte de consequente).
- **Levr:** representa a diferença entre dois itens presentes em conjunto em uma transação e os dois itens que aparecem independentemente (suporte*dados - suporte antecedente *consequente suporte/dados2)

Figura 4.9 - Regras de associação de produtos por loja e mês com confiança de 100%

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent	Consequent
0.987	1.000	0.987	1.013	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, ICED TEA MANGA CONTINENTE 1.5L=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.987	1.000	0.987	1.013	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.987	1.000	0.987	1.013	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, ICED TEA MANGA CONTINENTE 1.5L=1, LEITE UHT SEL M/G MATINAL 1LT=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.987	1.000	0.987	1.013	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, LEITE UHT SEL M/G MATINAL 1LT=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.987	1.000	0.987	1.013	1.000	0.000	ICED TEA MANGA CONTINENTE 1.5L=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT SEL M/G MATINAL 1LT=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.987	1.000	0.987	1.013	1.000	0.000	LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, LEITE UHT SEL M/G MATINAL 1LT=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.981	1.000	0.981	1.019	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT SEL M/G MATINAL 1LT=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.981	1.000	0.981	1.019	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE UHT SEL M/G MATINAL 1LT=1, LEITE UHT SEL M/G MATINAL 1LT=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.981	1.000	0.981	1.019	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, ICED TEA MANGA CONTINENTE 1.5L=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT SEL M/G MATINAL 1LT=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.981	1.000	0.981	1.019	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT SEL M/G MATINAL 1LT=1, LEITE UHT SEL M/G MATINAL 1LT=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.981	1.000	0.981	1.019	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, ICED TEA MANGA CONTINENTE 1.5L=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.981	1.000	0.981	1.019	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.981	1.000	0.981	1.019	1.000	0.000	ICED TEA MANGA CONTINENTE 1.5L=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.981	1.000	0.981	1.019	1.000	0.000	LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.981	1.000	0.981	1.019	1.000	0.000	ICED TEA MANGA CONTINENTE 1.5L=1, LEITE UHT SEL M/G MATINAL 1LT=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.981	1.000	0.981	1.019	1.000	0.000	LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, LEITE UHT SEL M/G MATINAL 1LT=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.981	1.000	0.981	1.019	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, ICED TEA MANGA CONTINENTE 1.5L=1, MANTEIGA MAGRA LIGHT MATINAL 250GR=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.981	1.000	0.981	1.019	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, MANTEIGA MAGRA LIGHT MATINAL 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.981	1.000	0.981	1.019	1.000	0.000	ICED TEA MANGA CONTINENTE 1.5L=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, MANTEIGA MAGRA LIGHT MATINAL 250GR=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.981	1.000	0.981	1.019	1.000	0.000	LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, MANTEIGA MAGRA LIGHT MATINAL 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.981	1.000	0.981	1.019	1.000	0.000	ICED TEA MANGA CONTINENTE 1.5L=1, LEITE UHT SEL M/G MATINAL 1LT=1, MANTEIGA MAGRA LIGHT MATINAL 250GR=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.981	1.000	0.981	1.019	1.000	0.000	LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, LEITE UHT SEL M/G MATINAL 1LT=1, MANTEIGA MAGRA LIGHT MATINAL 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.981	1.000	0.981	1.019	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, ICED TEA MANGA CONTINENTE 1.5L=1, MANTEIGA PAST C/SAL PRIMOR CLUVETE 250GR=1	-- CREME CULINARIO VAQUEIRO 250GR=1
0.981	1.000	0.981	1.019	1.000	0.006	ICED TEA MANGA CONTINENTE 1.5L=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, MANTEIGA PAST C/SAL PRIMOR CLUVETE 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.981	1.000	0.981	1.019	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, MANTEIGA PAST C/SAL PRIMOR CLUVETE 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, ICED TEA MANGA CONTINENTE 1.5L=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE UHT SEL M/G MATINAL 1LT=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE UHT SEL M/G MATINAL 1LT=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, ICED TEA MANGA CONTINENTE 1.5L=1, LEITE UHT SEL M/G MATINAL 1LT=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.975	1.000	0.975	1.026	1.000	0.000	LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT SEL M/G MATINAL 1LT=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.975	1.000	0.975	1.026	1.000	0.000	ICED TEA MANGA CONTINENTE 1.5L=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT SEL M/G MATINAL 1LT=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.975	1.000	0.975	1.026	1.000	0.000	LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, LEITE UHT SEL M/G MATINAL 1LT=1, MANTEIGA C/SAL MIMOSA 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, MANTEIGA MAGRA LIGHT MATINAL 250GR=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, ICED TEA MANGA CONTINENTE 1.5L=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, MANTEIGA MAGRA LIGHT MATINAL 250GR=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1, MANTEIGA MAGRA LIGHT MATINAL 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE UHT SEL M/G MATINAL 1LT=1, MANTEIGA MAGRA LIGHT MATINAL 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, ICED TEA MANGA CONTINENTE 1.5L=1, MANTEIGA MAGRA LIGHT MATINAL 250GR=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE C/CHOCOLATE MIMOSA 4*200ML=1, LEITE UHT SEL M/G MATINAL 1LT=1, MANTEIGA MAGRA LIGHT MATINAL 250GR=1	-- ICED TEA MANGA CONTINENTE 1.5L=1
0.975	1.000	0.975	1.026	1.000	0.000	CREME CULINARIO VAQUEIRO 250GR=1, LEITE UHT SEL M/G MATINAL 1LT=1, MANTEIGA MAGRA LIGHT MATINAL 250GR=1	-- LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT=1

Fonte: Software Orange Data Mining

venda. Esta linha de resultado da análise é suportada pelos valores de suporte (*Supp*) 0.975, cobertura (*Covr*) 0.994, força (*Strg*) 1.013, lift 1.000 e alavancagem (*Levr*) -0.000.

Estes dados, na visão de negócio, significam dizer que, na prática, os consumidores têm maior propensão a comprar os produtos da visão consequente, quando também compram os produtos da visão antecedente. Estes dados podem ser utilizados para criação de campanhas promocionais direcionadas, uma vez que existam dados de compras individuais dos clientes que possam ser cruzados e identificar aqueles clientes com maior propensão de comprar um segundo produto e aumentar o gasto médio deste na empresa.

5. DISCUSSÃO DOS RESULTADOS

5.1. Discussão dos resultados das estatísticas descritivas

Com o objetivo de descrever o comportamento das vendas dos produtos da empresa retalhista, foram, primeiramente, identificadas as principais estatísticas descritivas que são capazes de ilustrar o cenário comercial por público e por armazém. Através destas informações, observa-se comportamentos diferentes entre os públicos no que diz respeito às vendas executadas.

Neste sentido, foi identificado que no público de classe económica “baixa”, as vendas líquidas médias são maiores que as dos demais públicos e que este comportamento é comum na maioria das famílias de produtos. Estes resultados mostram que a empresa tem maior relevância para o público “baixa” em relação ao que vende, demonstrando maiores valores de vendas médias. A situação é totalmente contrária quanto ao público de classe económica “média”, que representa uma menor relevância levando em consideração os resultados deste público.

Diferente das vendas médias, em termos de vendas totais, as conclusões são que o público que mais representa as vendas totais líquidas, em todas as famílias de produto, é o público “Médio/baixo”. Ficando assim, o público “Médio”, como o quarto e último público em vendas totais.

A análise também foi executada na perspectiva das lojas, para identificar o comportamento das vendas por loja e entender se há diferenças relevantes entre estas que possam representar informações importantes.

O resultado desta análise mostra que a loja de Camarinha é a que tem maiores vendas médias líquidas, com médias superiores na grande maioria das famílias de produtos. Sendo assim, esta é a loja mais importante para a empresa uma vez que tem maior concentração de vendas médias e pode ser uma loja que merece foco por parte do retalhista em termos de representação de vendas. Também foi constatado que a loja Massamá é a loja menos importante em termos de vendas médias líquidas e, portanto, uma loja com oportunidades de crescimento, visto que todas as outras lojas possuem vendas médias superiores e podem servir de exemplo para alavancar as vendas desta loja.

Em termos de vendas totais, as conclusões são que na maioria das famílias de produtos é a loja de Camarinha com maiores vendas. A loja com menor representatividade em vendas totais e médias é a loja de Massamá, com menores valores nas duas variáveis.

5.2. Discussão dos resultados da árvore de decisão

A análise da árvore do público por classe social mostra que o público de classe económica “médio/baixo” representa uma maior parcela das vendas em quantidade, mostrando assim um público com maior concentração de vendas e com provável maior potencial de consumo, já que há maior probabilidade deste público comprar um produto.

Neste sentido, ainda a respeito do mesmo público, é possível identificar que, nas vendas em que são vendidos acima de 21 produtos, representam 40,4% em quantidade de produtos. Mostrando que é interessante, uma vez que este retalhista possua uma base de compras por clientes, identificar compras superiores a 21 produtos e tomar ações (promocionais ou não) direcionadas para este público, pois estes clientes têm maior probabilidade de consumir.

Também foi possível entender que para este público, as vendas acima de 45,58 € representam 41,3% das famílias de produtos, e, portanto, clientes que possuam este patamar de gastos podem ser potenciais consumidores para a maior parte das famílias de produto da empresa e impulsionar ações como aumento da cesta de produtos comprada por estes.

A última informação importante identificada nesta análise diz que o público que cumpre todos os comportamentos citados acima e quando a venda é maioritariamente para a família de produtos congelados, passa a ser o público de classe económica baixa. Esta informação é importante para entender que, em ações que envolvam esta famílias de produtos, é mais interessante manter o foco no público de classe económica baixa.

Uma vez que o público de classe económica baixa passa a ser relevante para as vendas do retalhista quando diz respeito às vendas das famílias de produtos congelados, foi executada uma árvore de decisão especificamente para este público e foram identificados alguns dados relevantes. Entre estes resultados, foi observado que este público detém de 28,8% das vendas em quantidade e que, os registos de vendas com mais de 21 produtos

para este público representam 31,9%, o que representa fortes indícios de propensão de compra maior para clientes que comprem mais que 21 produtos e que seria interessante considerá-los nas ações de vendas.

Continuando a análise da árvore de decisão para o público de classe económica baixa, observou-se que as vendas acima de 45,58 € representam 35,8% das famílias de produtos, e, portanto, clientes que possuam este patamar de gastos podem ser potenciais consumidores para a maior parte das famílias de produto da empresa e impulsionar ações como aumento da cesta de produtos comprada por estes. Este público obtém vendas maioritárias na família de congelados, sendo assim a família chave para ações direcionadas ao público.

5.3. Discussão das regras de associação de produtos por loja e mês

A aplicação das regras de associação mostra que existem produtos com probabilidade maior de serem comprados em conjunto, uma vez que nas vendas foram identificados no mesmo mês e loja, em uma proporção estatisticamente relevante. A partir desse cruzamento, o algoritmo permitiu identificar um produto com maior frequência nesta combinação, que configura assim um produto relevante para ações comerciais de aumento da quantidade de produtos comprados por cliente.

No exemplo da Figura 4.9, lê-se que quando os produtos CREME CULINARIO VAQUEIRO 250G, ICED TEA MANGA CONTINENTE 1.5L e LEITE C/ CHOCOLATE MIMOSA 4*200ML têm registo de vendas num mesmo mês e loja, é mais provável que o produto LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT também tenha registo de venda. Em termos de negócio, este último produto, o conseqüente, aparece, portanto, representando um produto chave para ser vendido em promoção a clientes que compraram o conjunto de produtos mencionados, os antecedentes.

Levando os mesmos critérios em consideração, no segundo exemplo, quando os produtos LEITE C/ CHOCOLATE MIMOSA 4*200ML, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT e MANTEIGA C/ SAL MIMOSA 250GR têm registo de vendas num mesmo mês e loja, é mais provável que o produto CREME CULINARIO VAQUEIRO 250GR também tenha registo de venda.

Os dois exemplos são justificados por parâmetros confiança (*Conf*), suporte (*Supp*), cobertura (*Covr*), força (*Strg*), Lift e alavancagem (*Levr*). Sendo que o primeiro exemplo carrega um nível de confiança de 100,0% e, o segundo, de 99,3%, a literatura abordada diz que quanto maior o nível de confiança, maior a probabilidade de as combinações acontecerem. Neste sentido, é recomendável que se utilizem de exemplos com maiores níveis de confiança para as ações comerciais de aumento da quantidade de produtos comprados por cliente, uma vez estas teriam maiores chances de se converterem em compra.

5.4. Implicações teóricas

Apesar de terem sido publicados alguns estudos que abordam diversas vertentes do retalho sendo estudadas através de técnicas de *data mining* (Miller, Smith, McIntyre e Achabal (2010); Kwak, Duvvuri e Russell (2015); Hekimoğlua et al. (2019); Sethuraman, Gázquez-Abad e Martínez-López (2022)), assim como foi mencionado neste estudo, pouco foi estudado sobre a utilização de *data mining* no âmbito da definição e otimização de sortido para lojas de retalho.

A revisão de literatura deste estudo permite ter acesso de forma rápida e sumarizada a uma base de conhecimento teórico construída por artigos científicos sobre os temas de *data mining* e gestão do sortido. Este estudo em específico permite melhor estudar o tópico no âmbito do retalho de proximidade. Para além de se tratar de uma análise empírica, traz, assim, embasamentos teóricos àquilo que foi estudado nos registos comerciais do retalhista e contribuindo para a academia com aspetos práticos de negócio observados nos dados analisados, tal como o trabalho de Hekimoğlua et al. (2019) e de Martinez et al. (2021) .

5.5. Implicações para a gestão de marketing

Rousseau (2014) defende que, quanto à previsão da composição de um sortido, pode-se definir de forma empírica ou por meio de planos de sortido rigorosos e que vão ao pormenor. O mesmo autor define que um plano de sortido como “a descrição pormenorizada das gamas a propor à clientela durante um determinado período” (Rosseau, 2014: 175). Contudo, é crucial e uma vantagem competitiva para o retalhista

entender quais são os produtos mais comprados pelos seus clientes e otimizar o número de produtos de acordo com a performance de cada um deles (Pratas e Brito, 2019).

Neste sentido, a contribuição destes estudos para o mercado, vai ao encontro do que foi citado pelos autores (Martinez *et al.*, 2021; Pratas e Brito, 2019), auxiliar ao retalhista objeto de estudo, com o entendimento do comportamento da venda de seus produtos e dos atributos que os constituem que são mais relevantes para cada tipo de público e loja. Estas informações são relevantes para apoiar nas decisões gerenciais, no ato de criação de ações promocionais, bem como ajustes na dimensão do sortido, uma vez que é conhecido aquilo que é mais importante em termos de vendas.

Para o setor retalhista, este estudo também tem o papel de discutir teorias e análises empíricas que podem ajudar os retalhistas interessados a unir informações relevantes sobre a gestão do sortido e como utilizar o *data mining* para criar análises que possam otimizá-lo, tal como estudado por Martinez *et al.* (2021).

6. CONCLUSÕES

6.1. Principais conclusões do estudo

O problema de investigação deste estudo que foi proposto, após revisão de literatura, onde foi identificado que há oportunidades de explorar o tema de otimização de sortido, mais especificamente utilizando técnicas de *data mining*. Neste sentido, foram definidos os objetivos de segmentar as lojas do retalhista, de acordo com a performance de vendas de seus produtos; identificar quais são os principais produtos vendidos para cada tipo de público; compreender quais são os principais produtos a serem considerados pelo retalhista para as lojas.

O estudo apresentou pontos de vista de autores que permite uma visão mais ampla sobre o novo contexto do retalho e o comportamento do consumidor no ato de escolha de sortido. A literatura discutiu a importância da gestão de sortido no retalho, com ênfase no formato de proximidade e como as mudanças no cenário atual do comportamento do consumidor aumenta a responsabilidade da gestão do sortido. Os autores citados também discutem a relevância da disponibilidade do sortido ideal e no espaço ideal, uma vez que tais processos são citados como uma vantagem competitiva para o negócio.

Entre os resultados obtidos, a literatura abordada diz que, de facto, o negócio de retalho sofreu grandes mudanças devido à mudança no comportamento do consumidor, o que aumentou a responsabilidade da gestão de sortido (Hekimoğlu *et al.*, 2019). Além disso, a gestão de sortido é não só uma ação concorrencial, mas sim um ato fundamental para o retalhista, uma vez que definir o que é oferecido para o consumidor envolve decisões de espaço e experiência no processo de compra (Rousseau, 2014).

Mais especificamente, para o retalho de proximidade, a literatura discute que o desafio de estabelecer uma estratégia a nível do sortido é ainda maior, pois o consumidor que se dirige para este formato, procura por conveniência, agilidade na compra e confiança (Gahinet e Cliquet, 2018). Assim, é crucial e uma vantagem competitiva para este tipo de retalhista entender quais são os produtos mais comprados pelos seus clientes e otimizar o número de produtos de acordo com a performance de cada um deles (Pratas e Brito, 2019).

Ainda no aspeto literário, recorreu-se a estudos onde a aplicação de *data mining* foi utilizada para estudar um problema específico do retalhista, envolvendo sortido e/ou estratégias para ações promocionais. A partir desta discussão da literatura, foi definida a metodologia para este estudo, onde foram utilizadas técnicas de *data mining*, nomeadamente árvore de decisão e regras de associação, a fim atender aos objetivos propostos para o estudo.

Por fim, no que diz respeito aos resultados obtidos na análise empírica da base de dados do retalhista, foram identificados resultados considerados relevantes, o que levou às conclusões descritas. Neste sentido, foram feitas análises descritivas, aplicação da árvore de decisão e da regra de associação.

Através das análises descritivas, foi possível concluir que o público “baixa” tem maior relevância nas vendas médias do retalhista e que, em termos de vendas totais, o público com maior relevância, em todas as famílias de produto, é o “Médio/baixo”. Assim, inferiu-se que essas duas classes económicas devem ser consideradas como principais nas decisões do retalhista.

As análises descritivas também foram aplicadas na perspectiva de loja, onde constatou-se que a loja de Camarinha tem maiores relevâncias em termos de vendas médias e totais. E, que a loja com menor relevância em vendas totais e médias é a loja de Massamá. Essas

informações são úteis para que o retalhista tenha a loja mais representativa (Camarinha) como foco nas suas decisões, e que utilize a loja de Massamá para identificar possíveis falhas para alavancar as vendas.

As conclusões retiradas da análise da árvore de decisão, mostram que há uma relevância com o público de classe “médio/baixo”, que tem uma maior relevância em quantidades. A análise também mostra atributos relevantes identificados nas vendas para este público, como o número de produtos maioritariamente identificado nas vendas e o valor de venda a partir do qual as vendas para este público é maioritária. Representando, portanto, uma informação a ser considerada para tomar ações (promocionais ou não) direcionadas para este público, pois estes clientes têm maior probabilidade de consumir.

A última informação importante identificada na análise de árvore de decisão diz que o público que cumpre todos os comportamentos citados acima e quando a venda é maioritariamente para a família de produtos congelados, passa a ser o público de classe económica baixa. Também foi identificado para este público informações relevantes como o número de produtos maioritariamente identificado nas vendas e o valor de venda a partir do qual as vendas para este público é maioritária, bem como a família de produtos em que as vendas do público prevalecem.

As conclusões retiradas da análise das regras de associação existem produtos com probabilidade maior de serem comprados em conjunto, uma vez que nas vendas foram identificados no mesmo mês e loja, em uma proporção estatisticamente relevante. Na conclusão, usou-se o exemplo que diz quem conforme diz a Figura 4.9, quando os produtos CREME CULINARIO VAQUEIRO 250G, ICED TEA MANGA CONTINENTE 1.5L e LEITE C/ CHOCOLATE MIMOSA 4*200ML têm registo de vendas num mesmo mês e loja, é mais provável que o produto LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT também tenha registo de venda.

Num segundo exemplo de regras de associação mostrou-se que quando os produtos LEITE C/ CHOCOLATE MIMOSA 4*200ML, LEITE UHT MAGRO MIMOSA BEM ESSENCIAL 1LT e MANTEIGA C/ SAL MIMOSA 250GR têm registo de vendas num mesmo mês e loja, é mais provável que o produto CREME CULINARIO VAQUEIRO 250GR também tenha registo de venda. Em termos de negócio nos dois exemplos, o último produto, o conseqüente, aparece, portanto, representando um produto chave para

ser vendido em promoção a clientes que compraram o conjunto de produtos mencionados, os antecedentes.

6.2. Limitações

A limitação mais relevante identificada neste estudo, foi em relação a incluir uma linha de contagem para entender se os públicos vão menos vezes comprar e assim, enriquecer as conclusões mencionadas neste estudo. Tal análise não é possível porque cada registo da base objeto de estudo não representa uma ocasião de venda e sim o total vendido naquele mês, a contagem aqui não representaria nada além disso. A mineração de dados de vendas ajuda a identificar comportamentos de compra do cliente, melhorar a qualidade do atendimento ao cliente, alcançar melhor retenção e satisfação do cliente, melhorar as taxas de consumo de mercadorias, adequar as políticas de transporte e distribuição de mercadorias mais pungentes e reduzir o custo dos processos de negócios (Martinez *et al.*, 2021).

6.3. Sugestões para estudos futuros

No mesmo sentido das limitações deste estudo, futuros estudos podem considerar informações a nível do cliente, por ocasião de compra, com o objetivo de extrair, através das mesmas técnicas utilizadas neste presente estudo (DA e AR), cruzamentos que comprovem quais são os atributos mais valorizados por tipo de cliente e quais os produtos vendidos em conjunto, neste caso com maior detalhamento, uma vez que será possível observar estas informações cliente a cliente.

Futuras pesquisas podem, inclusive, considerar famílias de produto diferentes das avaliada no presente estudo, como por exemplo, a família de produtos de eletrônicos, que possuem aspetos diferentes do consumo alimentar, onde podem ser extraídos comportamentos nas vendas e atributos a serem avaliados, contribuindo para um conhecimento do comportamento do consumidor em uma perspectiva complementar ao presente estudo.

7. REFERÊNCIAS

Agrawal, R., Aggarwal, C. & Prasad, V. (2000) Depth First Generation of Long Patterns. *Watson Research Center*, (1), 1- 11.

Arghode, V. (2012) Qualitative and Quantitative Research: Paradigmatic Differences. *Global Education Journal*, 4, 155- 163.

Bradlow, E., Gangwar, M., Kopalle, P. & Voleti, S. (2017) The Role of Big Data and Predictive Analytics in Retailing. *Journal of Retailing*, 93 (1), 79– 95.

Caigny, A., Coussement, K. & Bock, K. (2018) A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2), 760– 772.

Chen, Y., Chen, J. & Ching-Wen, T. (2006) A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales. *Decision Support Systems*, 42, 1503–1520.

Dogan, O., Ayçin, E. & Bulut, Z. (2018) Customer segmentation by using RFM model and clustering methods: a case study in retail industry. *International Journal of Contemporary Economics and Administrative Sciences*, 8 (1), 1-19

Gahinet, M. C. & Cliquet, G. (2018) Proximity and time in convenience store patronage: Kairos more than chronos. *Journal of Retailing and Consumer Services*, 43, 1-9.

Han, J., Pei, J., Yin, Y. & Mao, R. (2004) Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach. *Kluwer Academic Publishers*, 8 (1), 53– 87.

Hekimoğlua, M., Sevim, I., Aksezer A. & Durmuş, I. (2019) Assortment optimization with log-linear demand: Application at a Turkish grocery store. *Journal of Retailing and Consumer Services*, 50, 199– 214.

Jin, C., Li, F., Ma, S., & Wang, Y. (2022) Sampling scheme-based classification rule mining method using decision tree in big data environment. *Knowledge-Based Systems*, 244, 108522.

Kotler, P. & Armstrong, G. (2014) *Principals of Marketing*, 15th edition, Global Edition.

Krishna, G. & Ravi, V. (2020) Mining top high utility association rules using Binary Differential Evolution. *Engineering Applications of Artificial Intelligence*, 96, 103935. Doi: 10.1016/j.engappai.2020.103935 .

Kwak, K., Duvvuri, S. & Russell, G. (2015) An Analysis of Assortment Choice in Grocery Retailing. *Journal of Retailing*, 91 (1), 19– 33.

Martinez, M., Escobar B., García Diaz, M. & Pinto-Roa, D. (2021) Market basket analysis with association rules in the retail sector using Orange. Case Study: Appliances Sales Company. *CLEI Electronic Journal*, 24 (2), 1-17.

Miller C., Smith S., McIntyre S. & Achabal D. (2010) Optimizing and Evaluating Retail Assortments for Infrequently Purchased Products. *Journal of Retailing*, 86 (2), 159– 171.

Moodley R., Chiclana F., Caraffini F. & Carterb J. (2020) A product-centric data mining algorithm for targeted promotions. *Journal of Retailing and Consumer Services*, 54. Doi: 10.1016/j.jretconser.2019.101940

Neogrid (2018) Análise de sortimento: o que é e qual sua importância? Disponível em: <https://blog.neogrid.com/analise-de-sortimento-o-que-e-e-qual-sua-importancia/> (Acedido em 10 de julho de 2022).

Oliveira, P. (2000) *Tipos de Data Mining*. Disponível em: <https://paginas.fe.up.pt/~mgi99021/it/tipos.htm> (Acedido em 29 de agosto de 2022).

Orange (2022) Educação em Ciência de Dados. Disponível em: https://orangedatamining.com/home/teachers_and_students_love_it/ (Acedido em 28 de agosto de 2022).

Pratas J., & Brito, P. (2019) *Distribuição - Gestão de pontos de venda e de retalho*. Lisboa: Conjuntura Actual Editora, 37-388.

Rousseau, J. (2014) *Manual da Distribuição, 2ª edição*, Cascais: Príncípia Editora, 175-179.

Rygielski, C., Wang, C., & Yen, C. (2002) Data mining techniques for customer relationship management. *Technology in society*, 24 (4), 483-502.

Sethuraman, R., Gázquez-Abad, J. & Martínez-López, F. (2022) The effect of retail assortment size on perceptions, choice, and sales: Review and research directions. *Journal of Retailing*, 98 (1), 24– 45

Sharma, G. & Sharma P. (2016) A Study on Data Mining Algorithms for Tourism Industry. *International Journal of Latest Trends in Engineering and Technology*, 7 (1), 580-587.

Tekin, M., Etlioğlu, M., Koyuncuoğlu, Ö., & Tekin, E. (2018, August). Data mining in digital marketing. In *The International Symposium for Production Research*. Springer, Cham, 44-61.