



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Segmentação de clientes: um caso de aplicação em farmácia
comunitária

Ana Catarina Guerreiro Parrinha

Mestrado em Ciência de Dados,

Orientadora:

Doutora Patrícia Andreia da Silva Filipe, Professora Associada,
Iscte – Instituto Universitário de Lisboa

Setembro, 2024

Departamento de Métodos Quantitativos para Gestão e Economia
e Departamento de Ciências e Tecnologias da Informação

Segmentação de clientes: um caso de aplicação em farmácia
comunitária

Ana Catarina Guerreiro Parrinha

Mestrado em Ciência de Dados,

Orientadora:

Doutora Patrícia Andreia da Silva Filipe, Professora Associada,
Iscte – Instituto Universitário de Lisboa

Dedico esta dissertação aos meus pais.

Agradecimento

A conclusão desta dissertação representa o fim de uma etapa desafiadora e enriquecedora, pelo que me urge expressar a minha profunda gratidão a todos os que, de uma forma ou de outra, tornaram este percurso possível.

À minha orientadora, professora doutora Patrícia Filipe, agradeço por todo o apoio, a paciência e o conhecimento partilhado ao longo do projeto.

À empresa que forneceu a base de dados, reconheço o interesse e a disponibilidade demonstrados.

Aos meus amigos, tanto os que estão perto quanto os que estão longe, obrigada por toda a compreensão e o suporte manifestados durante o desenvolvimento deste trabalho.

À minha família, agradeço por todos os esforços efetuados em prol da realização deste mestrado, pela tolerância nos momentos em que não pude estar presente e pelo incentivo e confiança exteriorizados.

Resumo

A segmentação de clientes, ao facilitar a compreensão das necessidades dos compradores, constitui um procedimento fulcral para o sucesso das estratégias de *marketing*, sendo, por isso, adotada em variadíssimas áreas comerciais, como, nomeadamente, no campo farmacêutico.

O objetivo da dissertação consiste em auxiliar o setor de farmácia comunitária no seu negócio, ao fornecer divisões de mercado comportamentais, que permitam aos decisores obter *insights* valiosos acerca do seu público. A investigação concretiza-se através da exploração prática dos algoritmos CLARA (*Clustering for LARge Applications*) e DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) para aplicar análises de recência, frequência e valor monetário despendido (RFM) e das técnicas *K-means* e HAC (*Hierarchical Agglomerative Clustering*) para executar, com base na categorização dos produtos adquiridos, sistemas de *purchased products structure* (PPS), utilizando, para tal, os dados das transações de produtos não sujeitos a receita médica registadas numa farmácia portuguesa.

Os resultados do melhor modelo RFM obtido, decorrente do método CLARA, apontam para a existência de quatro grupos de clientes com diferentes graus de envolvimento com o negócio, enquanto os produtos do melhor modelo PPS conseguido, resultante do algoritmo *K-means*, sugerem a presença de doze conjuntos de compradores com necessidades distintas, designadamente de um agrupamento que adquire artigos pertencentes a diversas categorias e onze agregados que concentram, pelo menos, um terço dos seus gastos num único setor de comercialização do estabelecimento. O estudo constata, assim, que a segmentação comportamental, baseada nos sistemas mencionados, é adequável ao contexto de farmácia.

Palavras-chave: segmentação de clientes; farmácia comunitária; modelo RFM; modelo PPS

Abstract

Customer segmentation is a crucial process for the success of marketing strategies, as it enables the understanding of buyers' needs, being therefore applied across a variety of commercial sectors, including the pharmaceutical field.

The main goal of this dissertation is to assist the community pharmacy industry in its business by providing behavioural market segmentations, that allow decision-makers to gain valuable insights into their audience. The research is carried out through the practical exploration of the CLARA (Clustering for LARge Applications) and DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithms to apply recency, frequency and monetary value spent (RFM) analyses and the K-means and HAC (Hierarchical Agglomerative Clustering) techniques to implement, based on the categorization of acquired goods, purchased products structure (PPS) systems, leveraging data from records of over-the-counter transactions at a Portuguese pharmacy.

The results of the best RFM model, obtained through the CLARA approach, indicate the existence of four customer groups with different levels of engagement with the business, while the outcomes of the best PPS model, achieved utilizing the K-means algorithm, suggest the presence of twelve groups of buyers with distinct desires, including one group that purchases items from various categories and eleven clusters that concentrate, at least, a third of their spending on a single commercialization section of the establishment. The study thus demonstrates that behavioural segmentation, based on the mentioned systems, can be effectively applied to the pharmacy field.

Keywords: customer segmentation; community pharmacy; RFM model; PPS model

Índice Geral

Agradecimento	iii
Resumo	v
Abstract	vii
Índice de Quadros	xi
Índice de Figuras	xiii
Capítulo 1. Introdução	1
1.1. Contexto Científico	1
1.2. Contributo Científico	2
1.3. Estrutura da Dissertação	3
Capítulo 2. Revisão da Literatura	5
2.1. Técnicas para Segmentação de Clientes	5
2.2. Casos de Aplicação de Segmentação de Clientes	7
2.3. Casos de Aplicação de Segmentação de Clientes em Farmácia Comunitária	9
Capítulo 3. Metodologia	13
3.1. Compreensão do Negócio	16
3.2. Compreensão de Dados	17
3.3. Preparação de Dados	18
3.3.1. <i>Dataset</i> de Transações	18
3.3.2. <i>Dataset</i> de Clientes	24
3.4. Modelação	27
3.4.1. Análises RFM	27
3.4.1.1. CLARA	27
3.4.1.2. DBSCAN	28
3.4.2. Análises PPS	30
3.4.2.1. <i>K-means</i>	31
3.4.2.2. HAC	33
Capítulo 4. Resultados e discussão	37

4.1. Resultados	37
4.1.1. Análises RFM	37
4.1.2. Análises PPS	39
4.2. Discussão	42
Capítulo 5. Conclusões	45
Referências Bibliográficas	47
Anexo A	51

Índice de Quadros

QUADRO 3.1 – Tarefas e respectivos <i>outputs</i> das diferentes fases da metodologia CRISP-DM	14
QUADRO 3.2 – Descrição e natureza das variáveis consideradas pertinentes	17
QUADRO 3.3 – Medidas de estatística descritiva das variáveis “VENDAS_VALOR_CIVA”, “VENDAS_MARGEM” e “VENDAS_QUANTIDADE” no <i>dataset</i> original	23
QUADRO 3.4 – Medidas de estatística descritiva das variáveis “VENDAS_VALOR_CIVA”, “VENDAS_MARGEM” e “VENDAS_QUANTIDADE” no <i>dataset</i> pré-processado	24
QUADRO 3.5 – Medidas de estatística descritiva das variáveis “LUCRO_PERC”, “QUANTIDADE”, “GENERICOS_PERC”, “MEDICAMENTOS_PERC” e “IDADE”	27
QUADRO 3.6 – Média das variáveis de agrupamento por <i>cluster</i> (CLARA)	28
QUADRO 3.7 – Média das variáveis de agrupamento por <i>cluster</i> (DBSCAN)	30
QUADRO 3.8 – Média das variáveis de agrupamento por <i>cluster</i> (K-means) – <i>Clusters</i> 1 a 6	31
QUADRO 3.9 – Média das variáveis de agrupamento por <i>cluster</i> (K-means) – <i>Clusters</i> 7 a 12	32
QUADRO 3.10 – Média das variáveis de agrupamento por <i>cluster</i> (HAC) – <i>Clusters</i> 1 a 6	34
QUADRO 3.11 – Média das variáveis de agrupamento por <i>cluster</i> (HAC) – <i>Clusters</i> 7 a 12	35
QUADRO 3.12 – Média das variáveis de agrupamento por <i>cluster</i> (HAC) – <i>Clusters</i> 13 a 17	35
QUADRO 4.1 – Métricas de avaliação dos resultados dos modelos CLARA e DBSCAN	38
QUADRO 4.2 – Frequência relativa das categorias das variáveis “CLIENTE_TIPO_PESSOA”, “CLIENTE_SEXO” e “RESIDENTE” por <i>cluster</i> (CLARA)	39
QUADRO 4.3 – Média das variáveis “LUCRO_PERC”, “QUANTIDADE”, “GENERICOS_PERC”, “MEDICAMENTOS_PERC” e “IDADE” por <i>cluster</i> (CLARA)	39
QUADRO 4.4 – Métricas de avaliação dos resultados dos modelos K-means e HAC	40
QUADRO 4.5 – Frequência relativa das categorias das variáveis “CLIENTE_TIPO_PESSOA”, “CLIENTE_SEXO” e “RESIDENTE” por <i>cluster</i> (K-means)	41
QUADRO 4.6 – Média das variáveis “LUCRO_PERC”, “QUANTIDADE”, “GENERICOS_PERC”, “MEDICAMENTOS_PERC” e “IDADE” por <i>cluster</i> (K-means)	42

Índice de Figuras

FIGURA 2.1 – Esferas de categorização dos artigos científicos compilados aquando da revisão sistemática de literatura	5
FIGURA 3.1 – Metodologia CRISP-DM	13
FIGURA 3.2 – Transformação do objetivo e da definição de sucesso empresariais para terminologia de processamento de dados	17
FIGURA 3.3 – Número de clientes por número de utentes associados	19
FIGURA 3.4 – Número de produtos por mercado comercial	20
FIGURA 3.5 – Número de produtos por primeira divisão da Classificação Anatómica da EPHMRA	20
FIGURA 3.6 – Transformação da variável “CLIENTE_DATA_NASCIMENTO”	20
FIGURA 3.7 – Registos correspondentes a devoluções de produtos	21
FIGURA 3.8 – Registos correspondentes à compra e à posterior devolução	21
FIGURA 3.9 – Registos correspondentes às múltiplas compras e à posterior devolução	22
FIGURA 3.10 – Registos correspondentes à compra e à posterior devolução de um menor número de unidades de produto do que as inicialmente adquiridas	22
FIGURA 3.11 – Correção dos registos correspondentes às compras com posterior devolução de um menor número de unidades de produto do que as inicialmente adquiridas	22
FIGURA 3.12 – Transformação das variáveis “CLIENTE_FREGUESIA”, “CLIENTE_CONCELHO” e “CLIENTE_DISTRITO”	23
FIGURA 3.13 – Número de clientes por farmácia	24
FIGURA 3.14 – Distribuição da variável “RECENCIA”	25
FIGURA 3.15 – Distribuição da variável “FREQUENCIA”	25
FIGURA 3.16 – Distribuição da variável “MONETARIO”	25
FIGURA 3.17 – Método do cotovelo (CLARA)	28
FIGURA 3.18 – Distâncias ao quinto vizinho mais próximo ordenadas	30
FIGURA 3.19 – Método do cotovelo (K-means)	31
FIGURA 3.20 – Dendrograma	34
FIGURA A.21 – Distribuição da variável “Ajudas técnicas”	51
FIGURA A.22 – Distribuição da variável “Bebé e mamã”	51
FIGURA A.23 – Distribuição da variável “Contraceção e sexualidade”	51
FIGURA A.24 – Distribuição da variável “Cuidados específicos – olhos e ouvidos”	51

FIGURA A.25 – Distribuição da variável “Dermofarmácia, cosmética e acessórios”	51
FIGURA A.26 – Distribuição da variável “Higiene e cuidado oral”	51
FIGURA A.27 – Distribuição da variável “Homeopatia”	52
FIGURA A.28 – Distribuição da variável “Manipulação, soro fisiológico e soluções de irrigação”	52
FIGURA A.29 – Distribuição da variável “Medição de parâmetros e testes analíticos”	52
FIGURA A.30 – Distribuição da variável “Nutrição”	52
FIGURA A.31 – Distribuição da variável “Saúde animal”	52
FIGURA A.32 – Distribuição da variável “Sistema digestivo”	52
FIGURA A.33 – Distribuição da variável “Sistema nervoso e cessação tabágica”	53
FIGURA A.34 – Distribuição da variável “Sistema respiratório”	53
FIGURA A.35 – Distribuição da variável “Sistemas musculo-esquelético e circulatório”	53
FIGURA A.36 – Distribuição da variável “Solares”	53
FIGURA A.37 – Distribuição da variável “Suplementos alimentares”	53

CAPÍTULO 1

Introdução

A segmentação de mercado foi introduzida por Wendell Smith (1956), ao apresentar a distinção e a complementaridade entre a mesma e a diferenciação de produto, defendendo que a primeira se foca em conseguir identificar e alcançar grupos específicos de consumidores com recurso a estratégias de *marketing* personalizadas e a segunda procura oferecer produtos únicos e diferenciadores face aos distribuídos pelas empresas concorrentes. A técnica de segmentação de clientes é definida como a divisão do mercado em conjuntos de compradores com necessidades, características ou comportamentos distintos, que podem reagir de modo singular a diferentes campanhas publicitárias (Kotler & Armstrong, 1999). Desta forma, o processo referido torna-se uma componente fulcral no desenvolvimento dos projetos de *marketing* das empresas, ao permitir formular produtos e respetivas técnicas de venda para os seus clientes atuais (Ansoff, 1957) e almejar novos segmentos como público-alvo para a elaboração de artigos futuros (McDonald & Dunbar, 2004).

A tática de segmentação de mercado permite uma otimização da alocação dos recursos financeiros e humanos das empresas, ao possibilitar a identificação dos grupos mais lucrativos e mais promissores e o ajuste das técnicas de venda para atender aos interesses específicos de cada segmento (Cooil et al., 2008). Dada a relevância mencionada da divisão de mercado para o sucesso dos comércioos, esta é uma estratégia adotada em variadíssimas áreas de negócio, nomeadamente no setor farmacêutico, com um valor de mercado estimado em 1,6 biliões de dólares americanos em 2023, segundo o portal Statista (2024).

1.1. Contexto Científico

A literatura atual relativa aos géneros de segmentação de clientes existentes aponta para a presença de sete principais abordagens possíveis quanto ao tipo de dados tratados – a demográfica, que recorre a fatores socioeconómicos; a psicográfica, que faz uso dos atributos e das características psicológicas; a comportamental, que estuda as interações com a marca e hábitos de compra; a baseada no lucro, que avalia a importância dos consumidores para a empresa; a assente no valor percebido, que averigua o valor que os clientes atribuem à entidade; a geográfica, que recorre à localização; e a fundamentada no tipo de utilização, que analisa os padrões únicos de uso (Zhou et al., 2011; Koli et al., 2023). Todavia, a grande maioria dos casos de aplicação da estratégia em causa versa na perspetiva comportamental, exequível a partir do histórico de transações.

Como já referido, a estratégia comportamental é bastante documentada em casos de estudo, maioritariamente da área do retalho, onde, quer através de algoritmos de *clustering* ou de arquiteturas de *deep learning*, são contemplados sistemas de recência, frequência e valor monetário (RFM) e suas variantes, análises de *purchased products structure* (PPS), baseadas na categorização dos produtos adquiridos, e modelos *shopping mission* (SM), focados no motivo pelo qual os clientes visitam a loja, na tentativa de obter uma divisão de mercado que permita uma otimização dos resultados das campanhas de *marketing* aplicadas pelas empresas. Não obstante, a literatura existente relativa a casos de aplicação de segmentação de mercado de farmácia comunitária apenas tem recorrido, a partir de questionários aplicados aos seus consumidores, às abordagens psicográfica e de valor percebido, descurando completamente do histórico de transações do negócio e da informação que o mesmo fornece.

1.2. Contributo Científico

Dado o contexto científico descrito, o objetivo geral do presente estudo consiste em auxiliar o setor de farmácia comunitária no seu negócio, ao fornecer divisões de mercado comportamentais, que permitam aos decisores obter *insights* valiosos acerca do seu público. Os objetivos específicos prendem-se, numa primeira instância, com a análise crítica da literatura existente referente a técnicas para segmentação de mercado e a casos de aplicação das mesmas, quer em farmácia quer em outras áreas, e, seguidamente, com a exploração prática, em contexto de farmácia comunitária, dos algoritmos CLARA e DBSCAN para concretizar análises RFM, assentes na recência, na frequência e no valor monetário despendido, e das técnicas *K-means* e HAC para executar modelos PPS, baseados na categorização dos produtos adquiridos, utilizando, para tal, os dados das transações eletrónicas de produtos não sujeitos a receita médica registadas, entre janeiro de 2022 e maio de 2024, numa farmácia portuguesa, localizada no distrito de Santarém. Acresce, ainda, denotar que o desenvolvimento computacional do atual projeto se alicerça na linguagem de programação *Python*, com recurso ao ambiente de desenvolvimento *Jupyter Notebook*.

A relevância do atual projeto detém-se com a importância da abordagem comportamental da estratégia de *marketing* em debate, na medida em que a mesma possibilita que os gestores das farmácias estabeleçam planos de ação que atraem o seu público-alvo, identifiquem nichos de mercado, otimizem os resultados das campanhas publicitárias e os recursos financeiros utilizados nas mesmas, desenvolvam ofertas personalizadas para os diferentes segmentos detetados e aumentem a lealdade do consumidor e os lucros atingidos. Importa, também, referir que o valor da presente dissertação prende-se também com a contribuição para a investigação científica da temática em causa, dada a escassez de literatura, em contexto de farmácia comunitária, de casos de aplicação de segmentação comportamental de clientes.

1.3. Estrutura da Dissertação

O estudo é constituído pelo presente capítulo de introdução, seguindo-se o capítulo de revisão da literatura, onde é apresentada uma revisão sistemática da literatura alusiva às técnicas para segmentação de mercado e aos casos de aplicação das mesmas, quer em farmácia comunitária quer em outros setores, o capítulo de metodologia, onde são descritas as diferentes etapas do desenvolvimento do projeto técnico, como a compreensão do negócio, a compreensão dos dados, a preparação dos dados e a modelação, e o capítulo de resultados e discussão, onde são expostos e debatidos os produtos da dissertação. O último capítulo expõe as principais conclusões.

CAPÍTULO 2

Revisão da Literatura

O contexto científico do presente estudo delineou-se através da execução de uma revisão de literatura, incidente somente em artigos indexados nas bases de dados científicas *Web of Science* e *Scopus*, de modo a assegurar a qualidade e o rigor da mesma. Desta forma, para realizar a pesquisa de literatura, utilizou-se a *query* "customer segmentation" AND (("review" AND ("algorithm*" OR "method*" OR "technique*")) OR "transaction*" OR "purchas*" OR "pharmacy"), com o intuito de recolher artigos cujo desenvolvimento discorresse sobre técnicas para segmentação de clientes, oferecendo uma visão geral da utilização das mesmas, ou sobre casos de aplicação específicos destas. Sendo assim, após a obtenção dos 724 resultados correspondentes à *query* referida, dos quais 211 representavam documentos duplicados, 10 não dissertavam na língua inglesa e 107 não possuíam a expressão "customer segmentation" como palavra-chave, procedeu-se à averiguação da adequação dos restantes estudos aos objetivos pretendidos, através da verificação do título e do resumo dos mesmos. Optou-se, por fim, por considerar, para a presente revisão, 25 artigos científicos sintetizadores dos principais pontos detetados aquando da pesquisa de literatura, categorizando os mesmos, como se verifica na FIGURA 2.1, em três esferas distintas – técnicas para segmentação de clientes, casos de aplicação de segmentação de clientes e casos de aplicação de segmentação de clientes em farmácia comunitária.

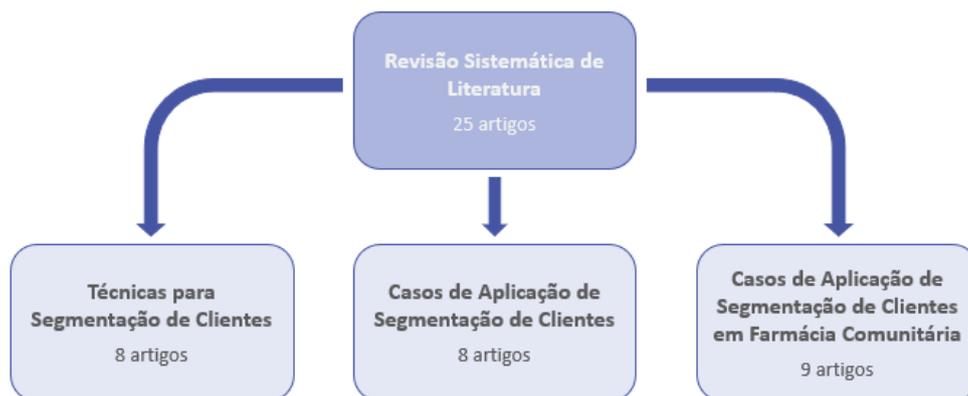


FIGURA 2.1. Esferas de categorização dos artigos científicos compilados aquando da revisão sistemática de literatura

2.1. Técnicas para Segmentação de Clientes

No que concerne aos géneros existentes de segmentação de clientes, Zhou et al. (2011) defendem a presença de cinco abordagens possíveis quanto ao tipo de dados utilizados – a demográfica, que

recorre a fatores sociodemográficos como idade, género, rendimento, nível de habilitações, religião e profissão; a psicográfica, que faz uso da personalidade, *hobbies*, estatuto social, opiniões, objetivos, valores, crenças e estilo de vida; a comportamental, que estuda os hábitos de consumo, as interações com a marca e a lealdade à mesma; a baseada no lucro, que avalia a importância dos consumidores para a empresa; a assente no valor percebido, que averigua o valor que os clientes atribuem à entidade. Já Koli et al. (2023) apresentam, para além das referidas segmentações demográfica, psicográfica e comportamental, uma abordagem geográfica, que recorre a atributos relativos à localização como clima, cultura, idioma e densidade populacional, e uma abordagem baseada no tipo de utilização, que analisa os padrões únicos de uso, distinguindo entre visitantes, visitantes com intenção de compra e compradores.

Relativamente à técnica utilizada para efetuar a segmentação de mercado, Sayan et al. (2022) afirmam que, aquando da seleção da mesma, não só deve ser considerada a dimensão da base de dados e a natureza da informação, como também as capacidades do algoritmo, para lidar com alta dimensionalidade e com dados com ruído, e a tipologia de *clusters* que o mesmo identifica. Os mesmos autores defendem, assim, que a escolha de um algoritmo deve ponderar o seu desempenho e a sua adequação à base de dados em causa. Sista et al. (2021) recomendam, ainda, a aplicação de *ensemble learning*, de modo a detetar qual a técnica que apresenta melhor desempenho aquando da segmentação de clientes.

Alusivamente aos estudos de segmentação de mercado existentes, Salminen et al. (2023) afirmam, através da análise de 172 artigos científicos, que os investigadores optam por utilizar só uma técnica, ao invés da combinação de duas ou mais, em 80% dos casos, que o *K-means* e as suas variantes são, sem dúvida, os algoritmos mais frequentemente utilizados, que a dimensão dos segmentos é o único hiperparâmetro definido em 81,7% dos artigos, que mais usualmente são sugeridos 4 grupos de clientes e que em apenas 4,1% dos estudos é solicitado o *feedback* de especialistas da área de aplicação em relação aos resultados obtidos, dependendo a validação do desempenho dos restantes projetos somente de métricas de avaliação. Quanto ao caso específico dos estudos de segmentação de mercado de negócios *online*, Gomes e Meisen (2023) constata, após a averiguação de 140 estudos, que o retalho é o caso de aplicação mais investigado, que o *K-means* é, mais uma vez, o algoritmo mais empregue e que em 52,4% dos artigos são utilizados métodos de seleção de variáveis, sendo que em 80% dos mesmos recorre-se à análise de recência, frequência e valor monetário (RFM). Importa, ainda, salientar que, apesar do modelo RFM ser bastante aplicado dado a sua simplicidade e facilidade de implementação e de compreensão, este não considera a informação pessoal e demográfica dos clientes, podendo, no entanto, ser alvo de modificações de modo a incorporá-la (Ernawati et al., 2021).

Por último, Hiziroglu (2013) dissertou sobre 42 artigos científicos de segmentação de mercado que recorreram a tecnologias *soft computing* – análises difusas, neuronais, evolutivas e aproximadas –, constatando que 65% dos mesmos utilizaram *neural computing* e que a técnica mais aplicada foram os *self-organizing maps* (SOM).

2.2. Casos de Aplicação de Segmentação de Clientes

O modelo RFM (recência, frequência e valor monetário) é uma das análises de segmentação de mercado mais difundidas, na medida em que possibilita uma divisão baseada no tipo de relação que os clientes estabelecem com o negócio. Doğan et al. (2018) aplicaram a técnica referida a uma cadeia de lojas de retalho de desporto turca, de modo a redefinir, através dos algoritmos de *clustering two-step* e *K-means*, os estatutos de cliente de bronze, de ouro e *premium*, previamente estabelecidos apenas de acordo com o fator monetário, sendo que o primeiro modelo corroborou a existência dos três segmentos de mercado já estipulados, mas com uma distribuição muito distinta, com 60% dos consumidores a serem realocados, e a segunda metodologia identificou, com valores sucessivamente melhores na análise RFM, quatro níveis de comprador – o “normal”, que integra os compradores ocasionais, correspondentes a 92% do mercado, o “avançado”, o “leal” e o “estrela”. Já Zhou et al. (2021), ao efetuarem a segmentação de mercado de uma loja de departamento americana com recurso a um método de *clustering* aglomerativo, optaram por introduzir um novo elemento T, medidor do intervalo temporal médio decorrido entre transações, ao modelo RFM, o que possibilitou a identificação de sete grupos de clientes, designadamente três segmentos já perdidos, um *cluster* em risco de abandono, uma classe promissora e dois *clusters* lucrativos e valiosos para a empresa. Por outro lado, Mahfuza et al. (2022), aquando da segmentação de mercado de uma *superstore* mundial, incorporaram os aspetos L e V na análise RFM, correspondentes, de modo respetivo, ao tempo passado entre a primeira e a última transação do cliente e à sua quantidade média de produtos adquiridos por dia de compras, com posterior redução de dimensionalidade do modelo em causa, através da análise de componentes principais (PCA), e divisão em *clusters*, com recurso aos algoritmos *K-means* e *K-medoids*, o que culminou na deteção de cinco segmentos de consumidores, nomeadamente dois grupos muito lucrativos, que devem ser retidos, dois agregados pouco rentáveis, que devem ser monitorizados, e um estrato que gera prejuízo financeiro à entidade, que deve ser substituído. Numa perspetiva distinta, Ho et al. (2023), a fim de efetuar a divisão do mercado *online* de uma loja americana através das técnicas *K-means* e *K-prototypes*, incluíram a componente D no modelo RFM, de forma a considerar características demográficas dos clientes, mais especificamente idade, região e género, o que resultou na identificação de cinco grupos de compradores – os leais, os relevantes para o negócio, os que necessitam de monitorização, os que estão em risco e os novos fregueses. Por último,

Liao et al. (2022), aquando da segmentação de clientes de uma plataforma chinesa de *e-commerce* através de um *self-organizing map* (SOM), introduzem uma alteração à esfera de comportamentos analisados pelo modelo RFM, ao considerarem, para os elementos relativos à recência e à frequência, todas as interações utilizador-item (clicar, marcar como favorito, adicionar ao carrinho e comprar), ao invés de atentarem apenas nas transações finalizadas, o que possibilitou a categorização dos consumidores em sete níveis – os essenciais, os preferenciais e os comuns para serem alvos de investimento, os regulares e os prioritários para serem focos de estratégias de retenção, os ocasionais e os dispensáveis.

Apesar de, como já mencionado, a análise RFM e as suas variantes serem um método de segmentação de mercado bastante comum, existem outras técnicas que permitem adquirir uma visão diferente para o mesmo procedimento. Deste modo, Sokol e Holý (2021) apresentam, numa aplicação a uma cadeia de drogarias checa, a comparação, a partir do algoritmo *K-means*, entre a análise de *purchased products structure* (PPS), baseada na categorização dos produtos adquiridos, e o modelo *shopping mission* (SM), focado no motivo pelo qual os clientes visitam a loja (compra de um artigo específico ou de bens variados), alegando que todas as vertentes devem ser utilizadas em conjunto, na medida em que cada uma fornece uma informação distinta. A aplicação do sistema PPS permitiu, entre as 55 tipologias existentes de produtos, onde 15 das mesmas totalizam mais de 98% das receitas da drogaria, a identificação de doze conjuntos de compradores, mais concretamente um agrupamento que adquire artigos pertencentes a diversas categorias e onze grupos cujas compras consistem maioritariamente num género específico de bens (artefactos de limpeza, detergentes para a roupa, produtos para o corpo, itens para a cara, produtos dentários, artigos para o cabelo, cosméticos, mercadoria para homem, bens para criança, perfumes ou itens sazonais). Já a execução do modelo SM foi finalizada em duas fases: a primeira, onde se efetuou o agrupamento dos cestos de compras baseado na categorização dos produtos neles presentes, o que gerou a deteção de doze *clusters*, com dois conjuntos a representarem cestos com bens variados, mas com faixas de preços distintas, e os restantes dez conglomerados a retratarem carrinhos maioritariamente carregados com um género de produtos (artefactos de limpeza, detergentes para a roupa, produtos para o corpo, itens para a cara, produtos dentários, artigos para o cabelo, cosméticos, mercadoria para homem, bens para criança ou artigos de higiene feminina); a segunda, onde se realizou a segmentação dos clientes baseada nos arquétipos de cesto definidos na primeira fase, o que possibilitou a determinação de dezoito grupos de clientes, designadamente cinco que adquirem carrinhos de artigos pertencentes a categorias diversas, mas com distribuições de dinheiro despendido distintas, dez cujas transações indiciam foco numa das tipologias de produtos identificadas na etapa anterior e três que se dirigem à loja tanto para a compra de cestos com produtos variados como também para a aquisição de uma única categoria de bens (artigos de limpeza, produtos para o cabelo ou cosméticos).

No que concerne à segmentação de clientes com recurso a *deep learning*, Nguyen (2021) discorre sobre a aplicação da estratégia referida a uma cadeia de supermercados retalhistas vietnamita através de uma arquitetura de rede neuronal, identificando, para cada transação, a categoria dos produtos comprados e a quantia alocada a cada uma das mesmas, o valor do item mais caro adquirido e o estatuto do cliente de acordo com o programa de lealdade da empresa. Este procedimento permitiu diferenciar os variados grupos de consumidores da entidade com base no seu comportamento de compras, o que motivou a definição de quatro grupos de clientes – os que maioritariamente compram alimentos frescos, especiarias e necessidades básicas, utilizando o estabelecimento como um substituto para as mercearias locais tradicionais; os que principalmente adquirem roupa e cosméticos; os que efetuam pequenas compras, dispondo do supermercado como uma loja de conveniência; os que sobretudo adquirem enlatados, alimentos processados, bebidas e doces.

Por fim, Nakano e Kondo (2018) apresentaram, a partir de uma análise de *clusters* de classes latentes e com recurso a dados obtidos através de uma empresa japonesa especializada em *marketing*, uma segmentação de mercado com base nos canais de compras utilizados (loja física ou *online*) e no tempo de uso de aparelhos eletrónicos pessoais, o que permitiu efetuar uma associação entre o meio de compras preferencial e o tempo despendido *online*. Deste modo, o estudo determinou a existência de sete tipos de compradores, designadamente quatro que optam por efetuar compras em loja, apesar de apenas dois dos mesmos manifestarem possuírem uma fraca presença *online*, um que demonstra ser pouco relevante para os comércios e dois que expressam interesse em dispor dos diferentes canais de compras, apresentando tempos de utilização de aparelhos eletrónicos acima da média.

2.3. Casos de Aplicação de Segmentação de Clientes em Farmácia

Comunitária

No que respeita à literatura relativa à segmentação de clientes de farmácia, existem variadíssimos estudos que apontam para uma divisão de mercado com base nos fatores que levam à escolha de determinado estabelecimento. Chereches et al. (2021), através de um questionário aplicado a uma amostra da população romena, identificam, após uma análise paralela para seleção de fatores, três segmentos de consumidores distintos – os interessados na farmácia em si e nas suas características como unidade de cuidados de saúde; os que valorizam o tipo de interação com o farmacêutico; e os “pragmáticos”, que, aquando da escolha de uma farmácia, apenas consideram os preços praticados e a proximidade ao seu domicílio. Já Lostakova et al. (2015) reconheceram, a partir de um inquérito a uma amostra da população checa, quatro grupos de clientes relativamente aos fatores considerados aquando da seleção de uma farmácia para a aquisição de medicamentos não sujeitos a receita médica e de suplementos alimentares – os despreziosos; os extremamente exigentes com tudo; os atentos

à qualidade do serviço e ao preço; e os particularmente sensíveis à qualidade do serviço. Alusivamente à segmentação de mercado de farmácias móveis, através de um questionário fornecido a uma amostra da população de uma cidade americana, Carroll e Gagon (1983) identificaram, após uma análise conjunta e ulterior aplicação de um método de *clustering* não hierárquico, cinco conjuntos de clientes distintos – os que valorizam a disponibilização de aconselhamento farmacêutico; os que apreciam a disponibilização de aconselhamento farmacêutico e a preservação dos seus registos de fármacos adquiridos; os que atribuem uma grande relevância aos preços praticados; os que prezam a simpatia dos funcionários; e os que sublimam o aconselhamento farmacêutico voluntário –, com posterior comparação da composição demográfica, do potencial de vendas, das despesas de farmácia e da lealdade à mesma entre os segmentos nomeados. Adicionalmente, Weilbaker et al. (1989) verificaram, no que respeita aos fatores valorizados e às fontes de informação utilizadas aquando da escolha de uma farmácia, as diferenças entre clientes de cadeias de farmácias e consumidores de farmácias independentes.

Apesar de, como já foi referido, a segmentação de clientes de farmácia comunitária baseada nos aspetos determinantes na escolha de um estabelecimento ser uma abordagem muito comum, existem estudos que, para além disso, também consideram, aquando da divisão de mercado, os fatores que influenciam a compra de medicamentos não sujeitos a receita médica. Deste modo, com recurso a um questionário fornecido a uma amostra da população de uma cidade grega, que visava avaliar os dois contextos anteriormente referidos, Kevrekidis (2018) reconheceu, através da aplicação do algoritmo de *clustering two-step*, a existência de três segmentos de clientes – os por conveniência; os leais; e os por conveniência e sensíveis aos preços praticados. Pribadi et al. (2020) efetuaram um estudo similar com o intuito de dividir os consumidores de farmácia indonésios.

Adotando uma outra perspetiva de segmentação, Schommer e Gaither (2014) optaram por agrupar os farmacêuticos e os utentes pelas suas posições relativamente ao papel do farmacêutico enquanto conselheiro da toma de medicação. Assim, após uma análise fatorial exploratória e posterior aplicação do algoritmo de *clustering two-step* às respostas a um inquérito relativo à temática explicitada, foram identificados quatro grupos de indivíduos – os que estão dispostos a abdicar do seu tempo para ouvir o aconselhamento farmacêutico, confiam no mesmo e não acreditam que devam existir restrições legais para tal; os que não querem dispensar tempo para escutar o aconselhamento farmacêutico; os que confiam exclusivamente nas recomendações médicas; e os que não pretendem prescindir do seu tempo para ouvir o aconselhamento farmacêutico, não dependem do mesmo e creem que devam existir regras legais para tal.

No que alude à segmentação de clientes de farmácias *online*, Patak et al. (2014), a partir de um questionário fornecido aos consumidores de uma e-farmácia checa, efetuaram uma divisão de mercado com base na importância atribuída às ferramentas de retenção, identificando, após uma análise fatorial e ulterior aplicação do algoritmo de *clustering two-step*, cinco conjuntos de utilizadores – os que valorizam as recompensas ganhas pela sua participação em inquéritos; os que apreciam o tipo de informação adquirida e os produtos grátis oferecidos; os que sublimam o grau de conhecimento obtido e os descontos cedidos; os que prezam os bens grátis e descontos oferecidos; os que não atentam em nenhum dos mecanismos de retenção, não os considerando como razão para voltar a realizar uma compra –, com posterior comparação da composição demográfica e comportamental, da escolha da e-farmácia e dos incentivos percebidos aquando da sua utilização entre os segmentos nomeados. Por outro lado, Gurău (2005), com base nas vantagens e desvantagens percebidas associadas às transações farmacêuticas *online*, identificou quatro tipos de consumidores de e-farmácia – os pouco interessados na qualidade do serviço e nos riscos do *online* e preocupados com os preços praticados e a segurança do pagamento; os que requerem privacidade e discrição aquando da entrega; os que exigem uma alta qualidade de serviço, anonimato e muitas opções de escolha e informação acerca das mesmas; os que valorizam a conveniência de comprar *online* e receber no domicílio e demonstram muita inquietação com os riscos das transações na *internet*.

Rematando, verificou-se que, apesar da segmentação de clientes ser um processo fundamental para o sucesso das estratégias de *marketing*, ao facilitar a compreensão das necessidades dos consumidores e, conseqüentemente, ao auxiliar as empresas a atender às demandas do mercado e a destacar-se da sua concorrência, a literatura científica existente relativa à divisão de mercado de farmácias comunitárias circunscreve-se, através de questionários aplicados ao seu público, à perspectiva psicográfica e à abordagem referente ao valor percebido pelos consumidores, descurando completamente do histórico de transações do negócio e da informação que este fornece. Deste modo, a presente dissertação visa abordar essa mesma lacuna, explorando, no contexto de farmácia comunitária, os modelos de segmentação comportamental RFM, assentes na recência, na frequência e no valor monetário despendido, e PPS, baseados na categorização dos produtos adquiridos, com recurso ao historial de compras dos clientes. Importa, ainda, ressaltar que, apesar da revisão de literatura efetuada versar em estudos presentes nas bases de dados científicas *Web of Science* e *Scopus*, foi, face a outras áreas de negócio, validada, através do *Google Scholar*, a escassez de artigos relativos à divisão comportamental de clientes no setor de farmácia.

CAPÍTULO 3

Metodologia

O desenvolvimento do presente estudo assentou na metodologia CRISP-DM (*Cross-Industry Standard Process for Data Mining*), constituída pelas seis fases visíveis na FIGURA 3.1, caracterizadoras do ciclo de vida de um projeto de ciência de dados. Chapman et al. (2000), aquando da introdução da metodologia referida como detalhada no QUADRO 3.1, apontaram a compreensão do negócio como a etapa inicial, onde são discutidos os objetivos e as condicionantes do projeto numa perspetiva empresarial e, de forma subsequente, definidos o problema de mineração de dados correspondente e o plano de trabalhos para o solucionar. Já a fase consecutiva do processo corresponde à compreensão dos dados, onde são efetuadas a recolha do *dataset* primitivo e a exploração inicial do mesmo, de modo a obter os primeiros *insights*, seguindo-se, assim, o ponto referente à preparação dos dados, onde é arquitetado o *dataset* final, através da seleção, limpeza e transformação da informação patente na base de dados original. O quarto passo da metodologia alude à modelação, onde são escolhidos e aplicados, com posterior otimização dos seus parâmetros, os algoritmos considerados pertinentes, de tal modo que a etapa seguinte corresponde à avaliação, onde é aferida a qualidade do modelo construído, o que compreende uma revisão de todas as práticas que levaram à sua conceção, e discutida a utilidade dos resultados obtidos. Por último, a sexta fase do método CRISP-DM equivale à implementação, onde o conhecimento adquirido é exposto à entidade interessada, quer seja através de um relatório final agregador de toda a informação relevante detetada ou através de um programa computacional desenvolvido para auxiliar os processos empresariais.

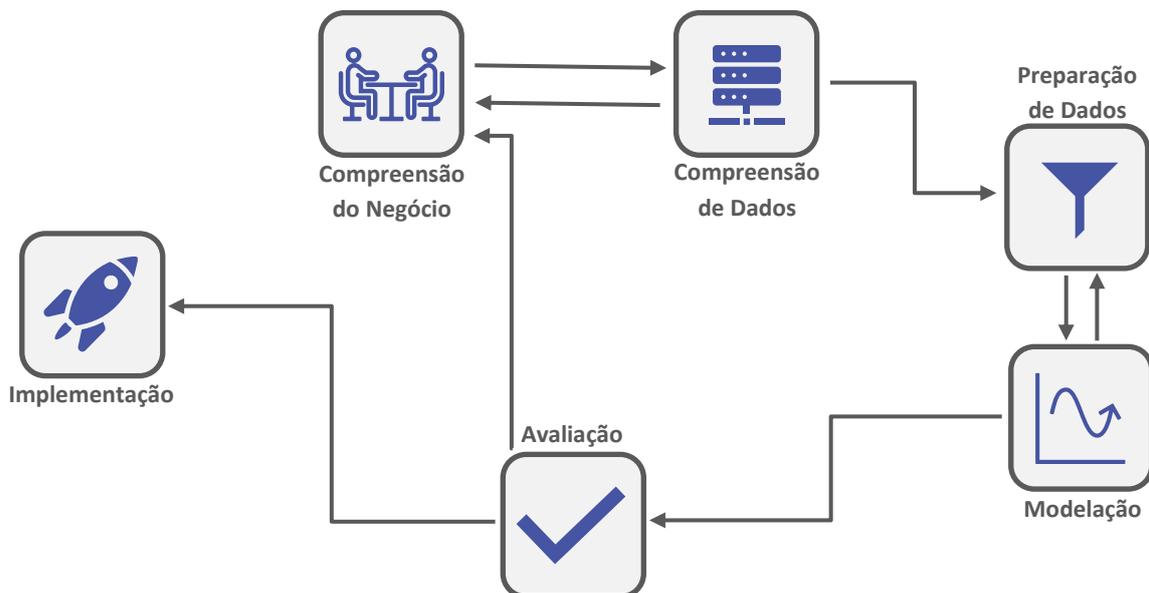


FIGURA 3.1. Metodologia CRISP-DM (adaptado de Chapman et al. (2000))

QUADRO 3.1. Tarefas e respetivos *outputs* das diferentes fases da metodologia CRISP-DM (adaptado de Chapman et al. (2000))

Fase	Tarefas	Outputs
Compreensão do Negócio	Determinar os objetivos do negócio	› <i>Background</i>
		› Objetivos do negócio
		› Critérios de sucesso definidos pelo negócio
	Aferir as circunstâncias do projeto	› Inventário de recursos
		› Exigências, premissas e limitações
› Eventuais constrangimentos e respetivos planos de contingência		
› Glossário de terminologia relevante		
Determinar os objetivos do processamento de dados	› Análise custo-benefício	
	› Objetivos do processamento de dados	
Compreensão de Dados	Efetuar o planeamento do projeto	› Critérios de sucesso técnico do processamento de dados
		› Plano do projeto
	Recolher os dados brutos	› Apreciação primitiva de ferramentas e técnicas
		› Relatório da recolha de dados brutos
		› Descrever os dados
› Relatório da descrição de dados		
Verificar a qualidade dos dados	› Relatório da exploração de dados	
	› Relatório de controlo de qualidade dos dados	
Preparação de Dados	Selecionar os dados apropriados	› Base lógica para inclusão e exclusão
	Limpar os dados	› Relatório da limpeza de dados
	Arquitectar dados	› Atributos derivados
		› Novas observações
	Integrar os dados	› Junção de bases de dados
	Formatar os dados	› Dados reformatados
		› Base de dados
› Descrição da base de dados		

QUADRO 3.1. Continuação

Fase	Tarefas	Outputs
Modelação	Selecionar a técnica de modelação	<ul style="list-style-type: none"> › Técnica de modelação › Pressupostos da técnica de modelação
	Elaborar o processo de teste	› Processo de teste
	Desenvolver o modelo	› Configurações dos parâmetros
		› Modelos
	Aferir o modelo	› Descrição do modelo
› Apreciação do modelo		
Avaliação	Avaliar os resultados	› Configurações dos parâmetros retificadas
		› Apreciação dos resultados do processamento de dados perante os critérios de sucesso definidos pelo negócio
	› Modelos aprovados	
	Rever o processo	› Revisão do processo
Implementação	Determinar os próximos passos	› Lista das ações possíveis
		› Decisão
	Planear a implementação	› Plano da implementação
	Planear a monitorização e a manutenção	› Plano de monitorização e manutenção
		Elaborar o relatório final
› Apresentação final		
Rever o projeto	› Registo das experiências significativas	

Note-se, ainda, que, apesar do QUADRO 3.1 destacar todos os procedimentos e respetivos *outputs* associados à metodologia CRISP-DM, a presente dissertação apenas visou aqueles considerados adequados às especificações do estudo em causa e ao contexto corrente de trabalho de investigação, descurando os mais oportunos em conjunturas de projetos empresariais.

3.1. Compreensão do Negócio

As entidades pertencentes à indústria farmacêutica, como as alusivas a tantas outras áreas, constataam a necessidade de se adaptarem às constantes transformações do mercado, de modo a suprirem eficazmente as vontades dos seus consumidores e, assim, se destacarem no paradigma competitivo empresarial. Nesse seguimento, o setor de farmácia comunitária apresenta como escopo a estratégia de *marketing* relativa à segmentação de clientes, que, através de uma divisão do mercado em conjuntos de compradores com necessidades, características ou comportamentos distintos, possibilita o desenvolvimento de planos de atração do seu público-alvo, a identificação de nichos de mercado, a otimização dos resultados das campanhas publicitárias e dos recursos financeiros alocados às mesmas, a elaboração de ofertas personalizadas para os diferentes segmentos detetados e o aumento da lealdade do consumidor e dos lucros atingidos. Sumariamente, na perspetiva do negócio, a utilidade do presente projeto prende-se com a aquisição de *insights* valiosos acerca dos seus consumidores, que possam, de alguma forma, auxiliar os decisores nas suas funções.

A questão empresarial anteriormente descrita traduz-se, em termos de processamento de dados, num problema de aprendizagem não supervisionada, cuja finalidade consiste em encontrar algum tipo de estrutura nos dados de *input*, designadamente agrupando os objetos que contenham um determinado grau de semelhança entre si. Deste modo, como visível na FIGURA 3.2, o objetivo técnico do estudo consiste, por meio dos dados transacionais de uma farmácia comunitária, na aplicação de algoritmos de *clustering*, nomeadamente dos métodos CLARA e DBSCAN para a execução de sistemas RFM, assentes na recência, na frequência e no valor monetário despendido pelos clientes, e das técnicas *K-means* e HAC para a concretização de análises PPS, baseadas na categorização dos produtos adquiridos. Por conseguinte, o sucesso técnico do projeto detém-se com a identificação de configurações de *clusters* apropriadas, com segmentos de mercado válidos e congruentes.

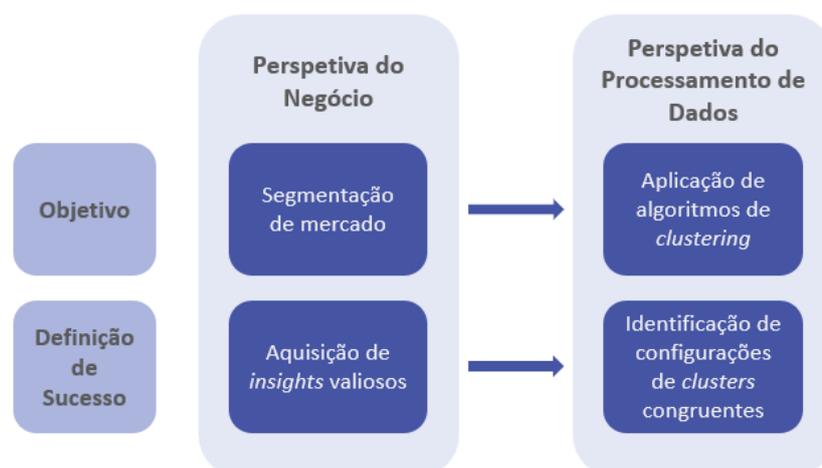


FIGURA 3.2. Transformação do objetivo e da definição de sucesso empresariais para terminologia de processamento de dados

3.2. Compreensão de Dados

O atual estudo dispôs, como base de trabalho, fornecida por uma empresa de consultoria especializada no ramo da saúde, de um *dataset* alusivo às transações de produtos não sujeitos a receita médica registadas, entre 1 de janeiro de 2022 e 22 de maio de 2024, em nove farmácias comunitárias portuguesas, dispersas pelos distritos de Lisboa, de Setúbal e de Santarém. O conjunto de dados, disponibilizado em formato *CSV*, contempla 1 321 660 observações e 100 atributos, compreendendo, assim, informações referentes aos aspetos das operações em si, às propriedades e à categorização dos produtos adquiridos e aos perfis dos compradores. Note-se que, visto o número elevado de variáveis cedidas, no QUADRO 3.2 apenas se encontram descritas aquelas consideradas relevantes para o escopo do presente projeto, com a sua posterior apreciação mais detalhada a ser apresentada aquando da etapa da preparação de dados.

QUADRO 3.2. Descrição e natureza das variáveis consideradas pertinentes

Variável	Descrição	Tipo
FARMACIA_NUM_TMP	Identificador da farmácia	Catégorico
CLIENTE_ID	Identificador do cliente	Catégorico
UTENTE_ID	Identificador do utente	Catégorico
DATA_ID	Data da transação	Data
VENDAS_VALOR_CIVA	Valor total da transação	Numérico
VENDAS_MARGEM	Valor da margem de lucro da transação	Numérico
VENDAS_QUANTIDADE	Número de unidades transacionadas	Numérico
PRODUTO_CODIGO	Identificador do produto	Catégorico
PRODUTO_DESIGNACAO	Nome do produto	Catégorico
PRODUTO_GENERICO	Tipologia da marca do produto	Catégorico
PRODUTO_GRANDE_MERCADO	Tipologia do produto	Catégorico
PRODUTO_DM_1_MERCADO	Mercado comercial ao qual o produto pertence	Catégorico
PRODUTO_DM_2_CATEGORIA	Categoria comercial à qual o produto pertence	Catégorico
PRODUTO_DM_3_SEGMENTO	Segmento comercial ao qual o produto pertence	Catégorico
PRODUTO_EPHMRA_N1	Primeira divisão à qual o produto pertence segundo a Classificação Anatómica da EPHMRA	Catégorico
PRODUTO_EPHMRA_N2	Segunda divisão à qual o produto pertence segundo a Classificação Anatómica da EPHMRA	Catégorico
PRODUTO_EPHMRA_N3	Terceira divisão à qual o produto pertence segundo a Classificação Anatómica da EPHMRA	Catégorico
PRODUTO_EPHMRA_N4	Quarta divisão à qual o produto pertence segundo a Classificação Anatómica da EPHMRA	Catégorico
UTENTE_COD_POSTAL	Código postal da morada fiscal do utente	Catégorico

QUADRO 3.2. Continuação

Variável	Descrição	Tipo
UTENTE_COD_POSTAL_LOCALIDADE	Localidade à qual pertence a morada fiscal do utente	Categórico
UTENTE_CONCELHO	Concelho ao qual pertence a morada fiscal do utente	Categórico
UTENTE_DATA_NASCIMENTO	Data de nascimento do utente	Numérico
UTENTE_DISTRITO	Distrito ao qual pertence a morada fiscal do utente	Categórico
UTENTE_FREGUESIA	Freguesia à qual pertence a morada fiscal do utente	Categórico
UTENTE_SEXO	Género do utente	Categórico
UTENTE_CARTAO_SAUDA_ASSOCIADO	Existência de vinculação ao Cartão Saúda do utente	Categórico
UTENTE_TIPO_PESSOA	Personalidade jurídica do utente	Categórico
CLIENTE_TIPO_PESSOA	Personalidade jurídica do cliente	Categórico
CLIENTE_SEXO	Género do cliente	Categórico
CLIENTE_DATA_NASCIMENTO	Data de nascimento do cliente	Numérico
CLIENTE_COD_POSTAL	Código postal da morada fiscal do cliente	Categórico
CLIENTE_FREGUESIA	Freguesia à qual pertence a morada fiscal do cliente	Categórico
CLIENTE_CONCELHO	Concelho ao qual pertence a morada fiscal do cliente	Categórico
CLIENTE_DISTRITO	Distrito ao qual pertence a morada fiscal do cliente	Categórico
CLIENTE_LATITUDE	Coordenada latitudinal da morada fiscal do cliente	Numérico
CLIENTE_LONGITUDE	Coordenada longitudinal da morada fiscal do cliente	Numérico

3.3. Preparação de Dados

A fase alusiva à preparação de dados dividiu-se em dois períodos, correspondendo o primeiro ao pré-processamento do *dataset* original, relativo às transações registadas, e o segundo à elaboração, a partir do *output* do ponto anterior, de uma base de dados focada nos consumidores.

3.3.1. *Dataset* de Transações

Visto a base de dados fornecida contemplar 100 atributos, o primeiro processo efetuado visou a seleção das variáveis relevantes para o presente projeto. Atentando no QUADRO 3.2, que destaca as colunas que se enquadram no escopo do estudo, procedeu-se às escolhas do utilizador da farmácia – “CLIENTE_ID” ou “UTENTE_ID” – e da categorização dos produtos – “PRODUTO_DM_1_MERCADO”, “PRODUTO_DM_2_CATEGORIA”, “PRODUTO_DM_3_SEGMENTO”, “PRODUTO_EPHMRA_N1”, “PRODUTO_EPHMRA_N2”, “PRODUTO_EPHMRA_N3” ou “PRODUTO_EPHMRA_N4” – a considerar. Relativamente ao utilizador da farmácia, o *dataset* compreendia 69 524 identificadores de clientes e 72 611 identificadores de utentes, com clientes a possuírem até 5 utentes associados, como se constata através da FIGURA 3.3, optando-se, deste modo, por reter os identificadores de clientes como

alvos para a segmentação de mercado, dado estes representarem os verdadeiros compradores do estabelecimento, ao contrário dos identificadores de utentes que, apesar de poderem corresponder ao consumidor final do produto, não são os clientes mais diretos do comércio. Já no que respeita à categorização dos produtos, prosseguiu-se, primeiramente, com a escolha da tipologia da mesma, aferindo-se que, segundo a classificação comercial, apenas 71 dos 13 630 produtos patentes na base dados não pertencem a nenhuma repartição, verificável a partir da FIGURA 3.4, e que, de acordo com a Classificação Anatómica da *European PHarmaceutical Market Research Association* (EPHMRA), 13 234 dos produtos não possuem rotulação, observável na FIGURA 3.5, pelo que, tendo em conta os objetivos da atual dissertação, avançar com a classificação comercial constituiu a escolha mais sensata. Considerando, então, apenas esta natureza de classificações, após a constatação de que os produtos se encontram divididos em 17 mercados, 77 categorias e 193 segmentos comerciais, optou-se por seguir com a catalogação em mercados comerciais, visto a mesma possibilitar uma visão mais ampla do negócio. Assim sendo, a seleção de atributos compreendeu as variáveis patentes no QUADRO 3.2, excetuando as colunas referentes aos utentes e aos seus perfis, à Classificação Anatómica da EPHMRA e às categorias e aos segmentos comerciais.

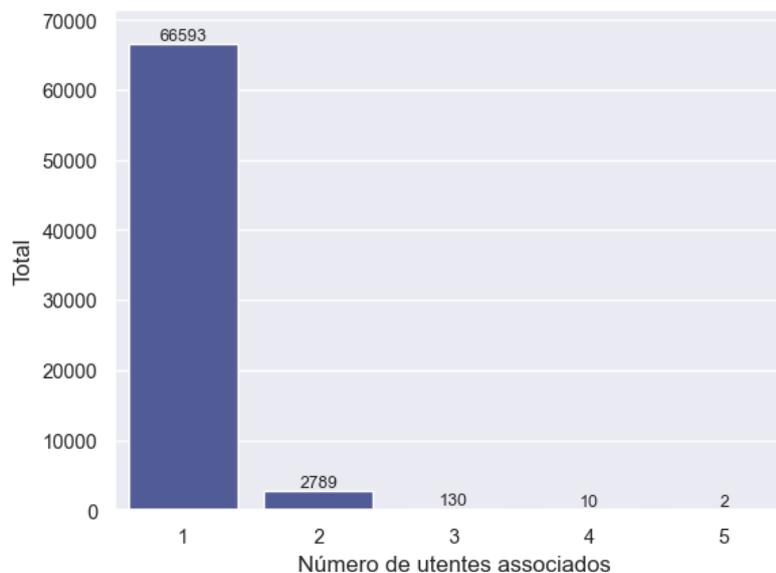


FIGURA 3.3. Número de clientes por número de utentes associados

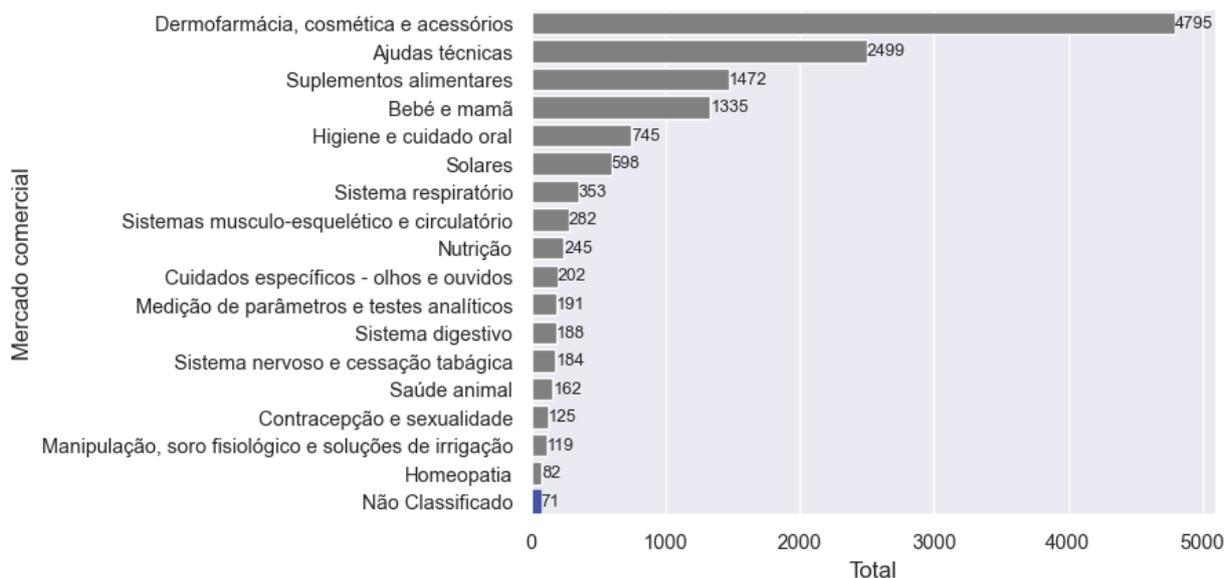


FIGURA 3.4. Número de produtos por mercado comercial

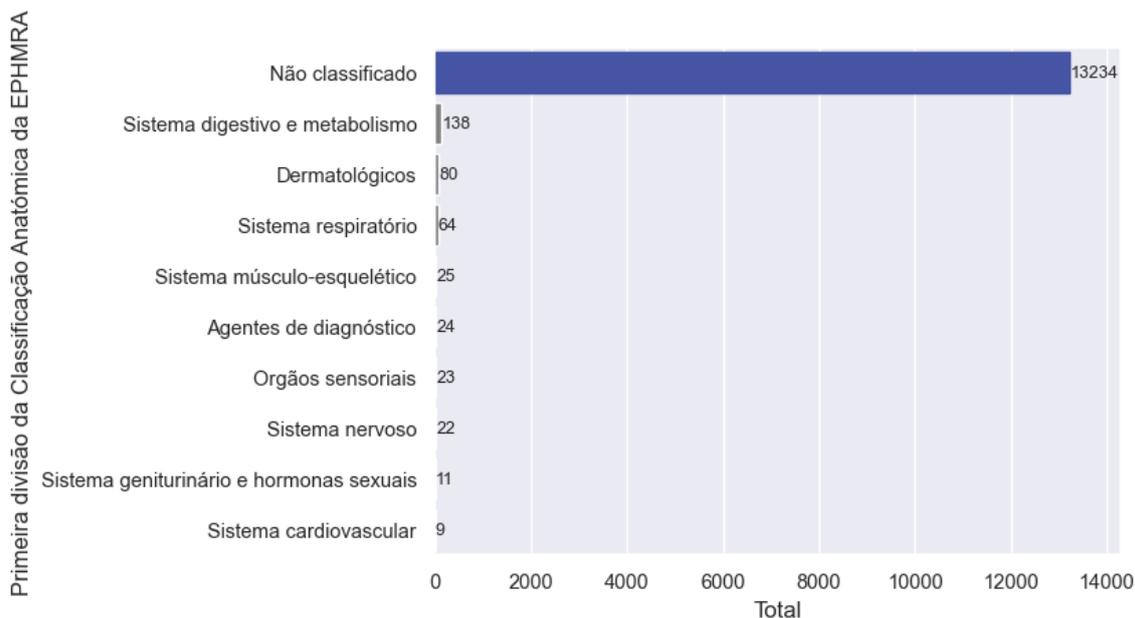


FIGURA 3.5. Número de produtos por primeira divisão da Classificação Anatômica da EPHMRA

De seguida, procedeu-se à correção do tipo dos atributos “FARMACIA_NUM_TMP”, “CLIENTE_ID”, “PRODUTO_CODIGO” e “DATA_ID”, definindo os três primeiros como variáveis categóricas e o último como coluna *datetime*, e à transformação da configuração e do tipo do atributo “CUSTOMER.BIRTHDATE” de formato numérico do *Excel* para o padrão ISO 8601 (AAAA-MM-DD), como exemplificado na FIGURA 3.6.



FIGURA 3.6. Transformação da variável “CLIENTE_DATA_NASCIMENTO”

Posteriormente, concretizou-se a uniformização dos valores nulos do *dataset*, estabelecendo a categoria “Não Classificado” da variável “PRODUTO_DM_1_MERCADO” e os códigos “0000-000” da coluna “CLIENTE_COD_POSTAL” como valores omissos. Posto isto, seguiu-se a eliminação das observações com informações essenciais em falta, nomeadamente com as variáveis “CLIENTE_ID” ou “PRODUTO_DM_1_MERCADO” como valores em branco, o que culminou na exclusão de 630 822 linhas.

No momento subsequente da preparação de dados, detetaram-se observações com valores negativos na variável “VENDAS_QUANTIDADE”, como as patentes na FIGURA 3.7, que foram interpretadas como devoluções de produtos e, por isso, eliminadas do *dataset*. Dada a remoção dessas linhas, considerou-se pertinente atuar também sobre os registos relativos às compras que precederam essas devoluções, visto que, para o objetivo da presente dissertação, um artigo retornado equivale a um artigo que nem chegou a ser adquirido. Nesse seguimento, foram identificados três tipos de situações distintas – a compra e a posterior devolução, como observável na FIGURA 3.8, o que suscitou a exclusão da linha relativa a essa venda; as múltiplas compras que culminam numa só devolução, como verificável na FIGURA 3.9, o que motivou a eliminação das observações alusivas a essas vendas; e a compra e a ulterior devolução de um menor número de unidades de produto do que as inicialmente adquiridas, como visível na FIGURA 3.10, o que promoveu a correção do registo referente a essa venda, para ajuste dos valores das variáveis “VENDAS_VALOR_CIVA”, “VENDAS_MARGEM” e “VENDAS_QUANTIDADE” para os equivalentes aos do número de unidades não retornadas, como exemplificado na FIGURA 3.11. Todo o processo descrito acarretou a remoção de 3 065 linhas da base de dados, correspondentes a devoluções e a compras que as antecederam.

	FARMACIA_NUM_TMP	CUSTOMER_ID	DATA_INTERNAL_ID	VENDAS_VALOR_CIVA	VENDAS_MARGEM	VENDAS_QUANTIDADE	PRODUTO_CODIGO
1679	14290	142900000018771.0	2024-03-14	-17.65	-7.93	-1	6201780
2455	14290	142900000023121.0	2023-08-28	-2.05	-0.96	-1	6127845
2791	14290	142900000020424.0	2023-12-14	-37.59	-0.02	-1	6472662

FIGURA 3.7. Registos correspondentes a devoluções de produtos

	FARMACIA_NUM_TMP	CUSTOMER_ID	DATA_INTERNAL_ID	VENDAS_VALOR_CIVA	VENDAS_MARGEM	VENDAS_QUANTIDADE	PRODUTO_CODIGO
1677	14290	142900000018771.0	2024-03-13	17.65	7.93	1	6201780
1679	14290	142900000018771.0	2024-03-14	-17.65	-7.93	-1	6201780

FIGURA 3.8. Registos correspondentes à compra e à posterior devolução

FARMACIA_NUM_TMP	CUSTOMER_ID	DATA_INTERNAL_ID	VENDAS_VALOR_CIVA	VENDAS_MARGEM	VENDAS_QUANTIDADE	PRODUTO_CODIGO	
527951	11614	116140000051271.0	2023-06-06	25.2	7.19	2	2443687
528006	11614	116140000051271.0	2023-06-07	12.6	3.33	1	2443687
527999	11614	116140000051271.0	2023-06-14	12.6	3.33	1	2443687
529476	11614	116140000051271.0	2023-06-25	-50.4	-13.31	-4	2443687

FIGURA 3.9. Registos correspondentes às múltiplas compras e à posterior devolução

FARMACIA_NUM_TMP	CUSTOMER_ID	DATA_INTERNAL_ID	VENDAS_VALOR_CIVA	VENDAS_MARGEM	VENDAS_QUANTIDADE	PRODUTO_CODIGO	
13154	14290	142900000042446.0	2022-03-04	13.50	3.58	2	2135184
13189	14290	142900000042446.0	2022-03-08	-6.75	-1.79	-1	2135184

FIGURA 3.10. Registos correspondentes à compra e à posterior devolução de um menor número de unidades de produto do que as inicialmente adquiridas

FARMACIA_NUM_TMP	CUSTOMER_ID	DATA_INTERNAL_ID	VENDAS_VALOR_CIVA	VENDAS_MARGEM	VENDAS_QUANTIDADE	PRODUTO_CODIGO	
13154	14290	142900000042446.0	2022-03-04	13.50	3.58	2	2135184
13189	14290	142900000042446.0	2022-03-08	-6.75	-1.79	-1	2135184



FARMACIA_NUM_TMP	CUSTOMER_ID	DATA_INTERNAL_ID	VENDAS_VALOR_CIVA	VENDAS_MARGEM	VENDAS_QUANTIDADE	PRODUTO_CODIGO	
13154	14290	142900000042446.0	2022-03-04	6.75	1.79	1	2135184

FIGURA 3.11. Correção dos registos correspondentes às compras com posterior devolução de um menor número de unidades de produto do que as inicialmente adquiridas

Seguidamente, procedeu-se à exclusão de observações incongruentes, designadamente daquelas com a variável “VENDAS_VALOR_CIVA” a apresentar valores negativos, com a coluna “VENDAS_QUANTIDADE” a igualar a zero ou com o atributo “VENDAS_MARGEM” a exibir um valor superior ao da variável “VENDAS_VALOR_CIVA”, o que conduziu à eliminação de 4 632 linhas. Note-se que registos com a coluna “VENDAS_VALOR_CIVA” a igualar a zero ou com a variável “VENDAS_MARGEM” a revelar valores negativos foram considerados coerentes, dado ser plausível um negócio oferecer artigos, por questões promocionais, e apresentar margens de lucro negativas em determinados produtos, por razões estratégicas.

O último procedimento efetuado correspondeu à regularização dos atributos referentes à residência fiscal dos clientes, como exemplificado na FIGURA 3.12, dado terem sido detetados valores em branco, erros e inconsistências na designação das localizações. Assim sendo, procedeu-se à substituição, de forma respetiva, dos valores das variáveis “CLIENTE_FREGUESIA”, “CLIENTE_CONCELHO” e “CLIENTE_DISTRITO” pelas freguesias, concelhos e distritos obtidos a partir do código postal patente na coluna “CLIENTE_COD_POSTAL” ou, caso o atributo referido constituísse um valor omissivo, do local resultante da junção da freguesia e do distrito disponíveis nas colunas “CLIENTE_FREGUESIA” e “CLIENTE_DISTRITO”, respetivamente.

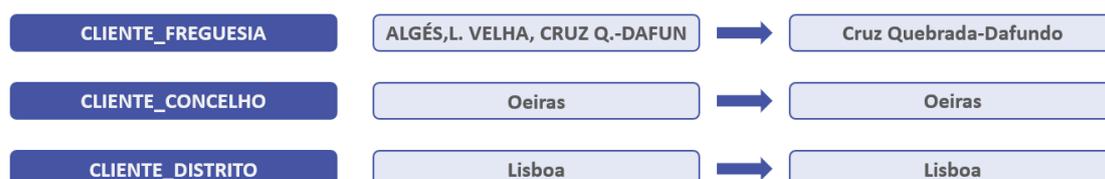


FIGURA 3.12. Transformação das variáveis “CLIENTE_FREGUESIA”, “CLIENTE_CONCELHO” e “CLIENTE_DISTRITO”

Findo o pré-processamento da base de dados original, relativa às transações registadas, foram retidas 683 141 observações, equivalentes a 51,69% do número de linhas iniciais. Nos QUADROS 3.3 e 3.4 estão presentes algumas medidas de estatísticas das variáveis referentes às vendas do *dataset* primordial e do *dataset* pós-preparação de dados, respetivamente.

QUADRO 3.3. Medidas de estatística descritiva das variáveis “VENDAS_VALOR_CIVA”, “VENDAS_MARGEM” e “VENDAS_QUANTIDADE” no *dataset* original

	VENDAS_VALOR_CIVA	VENDAS_MARGEM	VENDAS_QUANTIDADE
Mínimo	-7 709	-5 063,88	-650
Primeiro Quartil	6,75	2,26	1
Segundo Quartil	10,2	3,74	1
Terceiro Quartil	16,15	5,84	1
Máximo	14 300	107 321,52	1500
Média	13,51	5,02	1,33
Desvio-padrão	29,75	114,11	3,48

QUADRO 3.4. Medidas de estatística descritiva das variáveis “VENDAS_VALOR_CIVA”, “VENDAS_MARGEM” e “VENDAS_QUANTIDADE” no *dataset* pré-processado

	VENDAS_VALOR_CIVA	VENDAS_MARGEM	VENDAS_QUANTIDADE
Mínimo	0	-779,06	1
Primeiro Quartil	7,1	2,25	1
Segundo Quartil	10,5	3,65	1
Terceiro Quartil	16,55	5,63	1
Máximo	14 300	7 996	1500
Média	14,28	4,69	1,21
Desvio-padrão	34,25	15,39	3,16

Importa, ainda, referir que, aquando do início da presente etapa, ponderou-se reduzir a análise de divisão de mercado a compradores que constituíssem pessoas singulares, dispensando os que representassem pessoas coletivas, optando-se, no entanto, por englobar todos os tipos de clientes, visto essa inclusão permitir obter uma visão mais abrangente do mercado.

3.3.2. *Dataset* de Clientes

O primeiro procedimento efetuado, no começo da atual fase, visou, a partir do *dataset* de transações pré-processado obtido no ponto anterior, a elaboração de uma base de dados focada nos 67 736 consumidores identificados, abarcante das colunas referentes aos mesmos e aos seus perfis. Não obstante, de modo a alcançar uma segmentação de mercado mais precisa e significativa, optou-se, de seguida, por restringir o *dataset* a apenas uma farmácia, considerando, assim, somente os 17 273 clientes registados no estabelecimento “F”, localizado no distrito de Santarém, selecionado por ser aquele com mais compradores associados, como verificável na FIGURA 3.13.

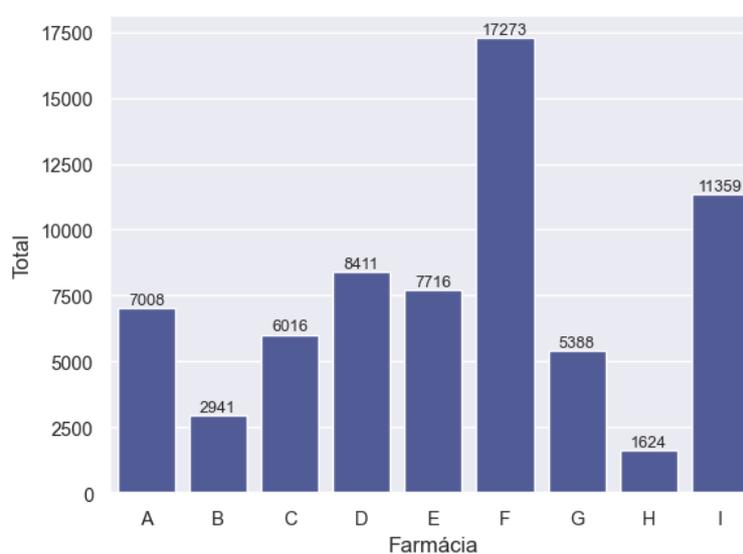


FIGURA 3.13. Número de clientes por farmácia

Posteriormente, de forma a possibilitar a exploração de sistemas RFM, assentes na recência, na frequência e no valor monetário despendido pelos clientes, realizou-se a estruturação das variáveis referentes ao tipo de relação que os compradores estabelecem com o negócio, designadamente do atributo “RECENCIA”, correspondente ao número de dias decorridos entre a última compra do utente e a finalização da recolha de dados de transações, da coluna “FREQUENCIA”, alusiva ao número de dias em que o consumidor visitou a farmácia, e da variável “MONETARIO”, relativa ao valor monetário despendido pelo comprador no comércio. As formas das distribuições dos atributos mencionados, patentes nas FIGURAS 3.14, 3.15 e 3.16, permitem constatar a presença de *outliers*, optando-se, todavia, pela retenção dos mesmos, com o intuito de capturar a totalidade do comportamento dos dados. Vista a fase de modelação subsequente, procedeu-se, ainda, à standardização das variáveis em causa, sendo este processo precedido pela aplicação da transformação Yeo-Johnson, de modo a obter distribuições mais aproximadas à gaussiana, com variâncias estabilizadas e assimetrias reduzidas, através da regra visível na FÓRMULA 3.1, onde x_i corresponde ao valor de cada observação, com $i = 1, 2, \dots, n$ e n a igualar a dimensão da amostra, e λ é estimado, para cada variável, através do método da máxima verosimilhança, de forma a minimizar a assimetria. Importa, assim, referir que as técnicas citadas corrigiram as questões pretendidas quanto às distribuições dos dados.

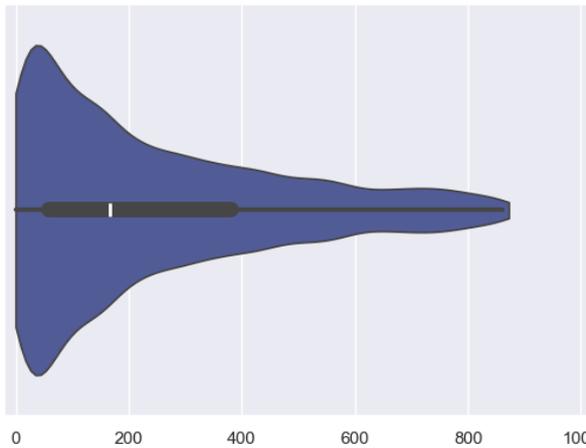


FIGURA 3.14. Distribuição da variável “RECENCIA”

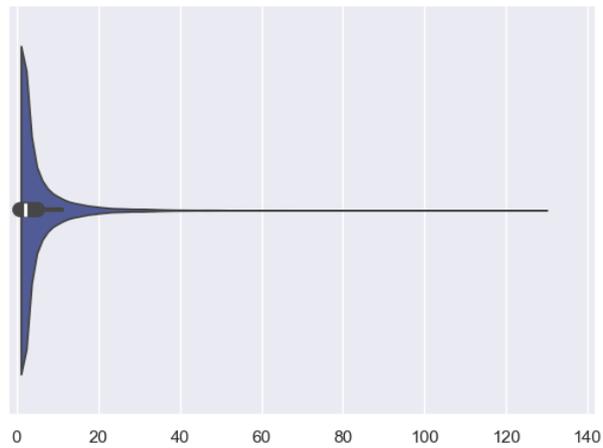


FIGURA 3.15. Distribuição da variável “FREQUENCIA”

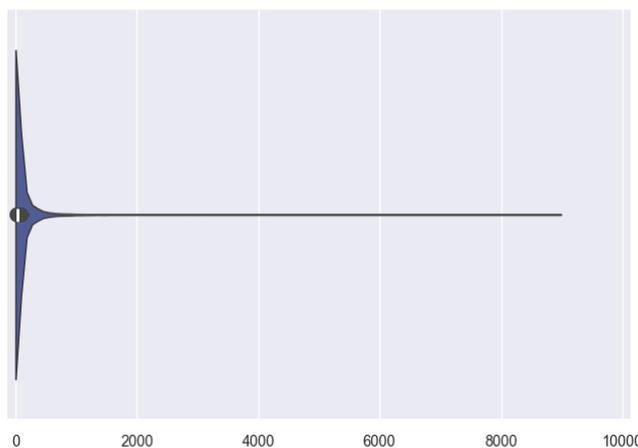


FIGURA 3.16. Distribuição da variável “MONETARIO”

$$x_i^{(\lambda)} = \begin{cases} [(x_i + 1)^\lambda - 1]/\lambda, & \lambda \neq 0 \text{ e } x_i \geq 0 \\ \ln(x_i + 1), & \lambda = 0 \text{ e } x_i \geq 0 \\ -[(-x_i + 1)^{2-\lambda} - 1]/(2 - \lambda), & \lambda \neq 2 \text{ e } x_i < 0 \\ -\ln(-x_i + 1), & \lambda = 2 \text{ e } x_i < 0 \end{cases} \quad (3.1)$$

De seguida, com o intuito de viabilizar a utilização de análises de *purchased products structure* (PPS), baseadas na categorização dos produtos adquiridos, efetuou-se a formulação dos atributos requeridos pelas mesmas, obtendo, para cada comprador, o rácio das suas despesas em cada um dos 17 mercados comerciais, o que permitiu apurar, através das distribuições patentes no Anexo A, que as divisões “Dermofarmácia, cosmética e acessórios”, “Sistema digestivo”, “Sistema nervoso e cessação tabágica”, “Sistema respiratório” e “Suplementos alimentares” são as que registam transações de um maior número de clientes, sendo que nenhuma das 17 esferas reúne compras de mais de 50% dos consumidores. Analogamente ao procedimento empregue às variáveis relativas ao tipo de relação que os clientes estabelecem com o negócio, prosseguiu-se, mais uma vez, com a aplicação da transformação Yeo-Johnson às novas colunas, de modo a reduzir a assimetria registada, e com a estandardização das mesmas, através da divisão pelo desvio-padrão respetivo, de forma a, assim, preservar a esparsidade dos dados. Acresce, assim, mencionar que os procedimentos referidos atenuaram, com sucesso, as questões pretendidas quanto às distribuições dos atributos.

Por último, procedeu-se à construção de variáveis adicionais, caracterizadoras dos clientes e do seu comportamento, designadamente do atributo “LUCRO_PERC”, relativo à percentagem de lucro média das transações do comprador, da coluna “QUANTIDADE”, alusiva ao número de itens adquiridos, da variável “GENERICOS_PERC”, respeitante à percentagem de fármacos genéricos comprados, face ao total de remédios adquiridos, do atributo “MEDICAMENTOS_PERC”, correspondente à percentagem de medicamentos comprados, perante o somatório de itens adquiridos, da coluna “IDADE”, equivalente à idade do consumidor, e da variável “RESIDENTE”, com a indicação de se a morada fiscal do cliente pertence ao mesmo distrito onde a farmácia se localiza. No QUADRO 3.5 estão patentes algumas medidas de estatística descritiva das variáveis numéricas apresentadas, sendo possível destacar que, em termos médios, as transações dos fregueses evidenciam uma taxa de lucro igual a 35,57%, com 56,54% dos 8 bens adquiridos a consistirem em medicamentos, dos quais apenas 7,7% correspondem a genéricos.

QUADRO 3.5. Medidas de estatística descritiva das variáveis “LUCRO_PERC”, “QUANTIDADE”, “GENERICOS_PERC”, “MEDICAMENTOS_PERC” e “IDADE”

	LUCRO_PERC	QUANTIDADE	GENERICOS_PERC	MEDICAMENTOS_PERC	IDADE
Mínimo	-1571,28	1	0	0	0
Primeiro Quartil	29,87	2	0	27,27	36
Segundo Quartil	35,26	3	0	57,14	51
Terceiro Quartil	40,82	8	0	100	66
Máximo	100	793	100	100	101
Média	35,57	8,04	7,7	56,54	50,34
Desvio-padrão	16,19	15,89	20,29	36,53	21,25

3.4. Modelação

A etapa referente à modelação de dados compreendeu duas partes, consistindo a primeira na aplicação, através dos algoritmos CLARA (*Clustering for LARge Applications*) e DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), de sistemas RFM, baseados na recência, na frequência e no valor monetário despendido pelos clientes, e a segunda na exploração, com recurso aos métodos K-means e HAC (*Hierarchical Agglomerative Clustering*), de análises PPS, assentes na categorização dos produtos adquiridos.

3.4.1. Análises RFM

De modo a prosseguir com a execução dos modelos RFM, alicerçados na recência, na frequência e no valor monetário despendido pelos consumidores, optou-se pela utilização de métodos de agrupamento mais apropriados para lidar com dados com *outliers*, tendo, assim, em particular consideração as distribuições das variáveis “FREQUENCY” e “MONETARY” patentes nas FIGURAS 3.15 e 3.16, respetivamente. Posto isto, selecionaram-se os métodos CLARA (*Clustering for LARge Applications*), por depender de medoides ao invés de centroides, o que torna o algoritmo mais robusto a casos extremos (Kaufman & Rousseeuw, 1990), e DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), por permitir a deteção de *outliers*.

3.4.1.1. CLARA

Segundo Qian et al. (2018), o algoritmo CLARA (*Clustering for LARge Applications*) consiste numa extensão do método PAM (*Partitioning Around Medoids*) para grandes bases de dados, ao utilizar apenas amostras aleatórias para aplicação da técnica, ao invés do *dataset* completo. O método começa por selecionar um subconjunto de dados e aplicar o algoritmo PAM ao mesmo, de forma a obter os

seus medoides, prosseguindo, posteriormente, com a associação de cada objeto do conjunto de dados completo ao seu medoide mais próximo e com o cálculo da dissimilaridade média do *clustering* obtido. Todo o processo descrito é feito na iteração seguinte, sendo que a repetição com o menor valor de dissimilaridade média é fixada como aquela cujos medoides oferecem o melhor agrupamento.

De modo a proceder com a execução da técnica CLARA na presente dissertação, à semelhança do estudo de Qian et al. (2018), definiram-se o tamanho dos subconjuntos de dados como 2% do número de clientes (350 objetos) e a quantidade de iterações em 10. Já seguindo Enesi et al. (2021), que sugere o método do cotovelo para a seleção do número ótimo de segmentos, o total de *clusters* foi estipulado em 4, como indica o ponto de inflexão visível na FIGURA 3.17. No QUADRO 3.6 estão patentes as principais características dos *clusters* obtidos com a análise RFM efetuada através do algoritmo CLARA, aplicado por meio da distância euclidiana, sendo possível constatar a existência de um conjunto promissor (*Cluster 1*), um grupo crucial para o faturamento do estabelecimento (*Cluster 2*) e dois agrupamentos que demonstram perda de interesse no negócio (*Clusters 3 e 4*).

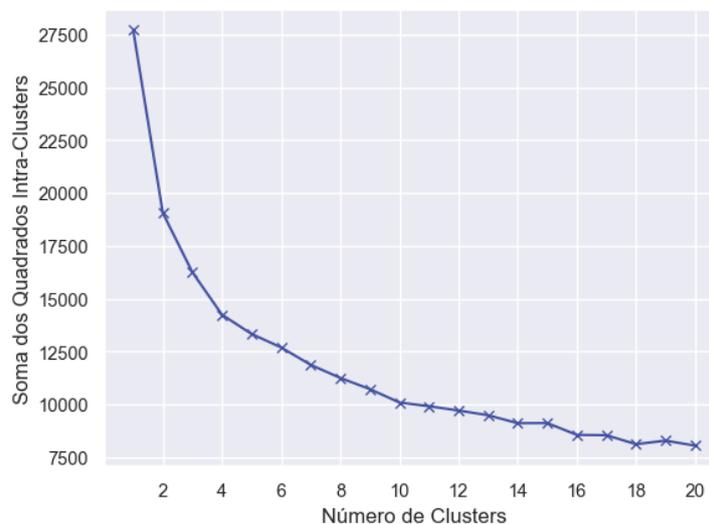


FIGURA 3.17. Método do cotovelo (CLARA)

QUADRO 3.6. Média das variáveis de agrupamento por *cluster* (CLARA)

	Cluster 1 19,75%	Cluster 2 25,22%	Cluster 3 33,24%	Cluster 4 21,79%
RECENCIA	66,05	63,47	416,49	356,15
FREQUENCIA	2,07	12,74	1,06	3,41
MONETARIO	33,75	281,05	15,4	63,71

3.4.1.2. DBSCAN

Ester et al. (1996) introduziram o método de segmentação DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) como a identificação de regiões com alta densidade de pontos no espaço

n -dimensional, separadas por zonas pouco densas. O algoritmo requer a definição de dois parâmetros essenciais, designadamente do raio Eps (épsilon), que determina a distância máxima a que dois pontos se podem encontrar para que sejam considerados vizinhança, e do valor $MinPts$, que corresponde ao número mínimo de objetos necessário para formar um *cluster*. Considerando os pontos p e q , a técnica em causa exige ainda a compreensão dos conceitos de vizinhança $N_{Eps}(p)$, onde q integra a vizinhança de p caso a distância entre os dois seja inferior ou igual ao parâmetro Eps , de ponto p diretamente alcançável por densidade por q , caracterizado por p pertencer à vizinhança de q e esta possuir um número de objetos igual ou superior ao valor da propriedade $MinPts$, de ponto p alcançável por densidade por q , definido por existir uma cadeia de pontos p_1, p_2, \dots, p_n , onde $p_1 = q$ e $p_n = p$, de forma a que p_{i+1} seja diretamente alcançável por densidade por p_i , com $i = 1, 2, \dots, n - 1$ e n a igualar a dimensão da cadeia, e de ponto p conectado por densidade a q , determinado por existir um objeto o , de modo a que ambos p e q sejam alcançáveis por densidade por o . Posto isto, o domínio dos conceitos apresentados permite, então, definir os três tipos de pontos identificados pelo método DBSCAN – os centrais, cuja vizinhança- Eps possui um número de objetos igual ou superior ao valor do parâmetro $MinPts$; os de fronteira, que são alcançáveis por densidade por pontos centrais; e os *outliers*, que não cumprem nenhuma das condições anteriormente enunciadas.

Tendo em consideração os conceitos apresentados, os *clusters* formados pelo algoritmo DBSCAN consistem em conjuntos de pontos conectados por densidade, com o máximo alcance por densidade atingível (Zakrzewska & Murlewski, 2005). De acordo com Birant e Kult (2007), a execução do método em estudo inicia-se com um ponto arbitrário p , selecionando, para um determinado raio Eps e valor $MinPts$, todos os objetos alcançáveis por densidade pelo mesmo. Caso p constitua um ponto central, forma-se um novo *cluster* com o mesmo e a respetiva vizinhança, sendo, seguidamente, calculados, de modo recursivo, todos os objetos conectados por densidade a qualquer ponto central desse agrupamento, com aqueles cujo número de objetos vizinhos seja inferior ao parâmetro $MinPts$ a serem classificados como *outliers*. A técnica descrita repete-se até que todos os pontos sejam processados.

De forma a prosseguir com a aplicação do algoritmo DBSCAN, por meio da distância euclidiana, na presente investigação, os valores dos parâmetros $MinPts$ e Eps estabeleceram-se conforme o indicado no estudo de Schubert et al. (2017), que aponta que o fator $MinPts$ deve corresponder ao dobro do número de dimensões da análise – $2 \times dim$ – e que a propriedade Eps deve ser definida com base na distância ao $(2 \times dim - 1)$ vizinho mais próximo. Assim sendo, a característica $MinPts$ foi fixada em 6, visto que o sistema RFM comporta 3 dimensões, e o parâmetro Eps foi estipulado em 0,24, após experiências com valores oscilantes entre 0,15 e 0,3, deduzidos de acordo com o cotovelo patente na FIGURA 3.18. No QUADRO 3.7 estão disponíveis os principais atributos dos 4 agrupamentos atingidos com a análise RFM efetuada através do método DBSCAN, onde apenas 205 casos foram apontados como *outliers*, sendo, então, possível apontar a existência de um grupo crucial para o faturamento do

estabelecimento (*Cluster 1*) e três conjuntos que demonstram perda de interesse no negócio (*Clusters 2, 3 e 4*).

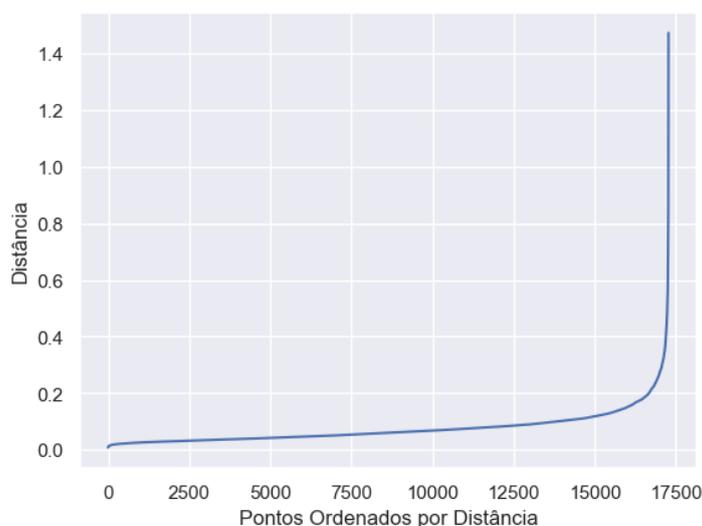


FIGURA 3.18. Distâncias ao quinto vizinho mais próximo ordenadas

QUADRO 3.7. Média das variáveis de agrupamento por *cluster* (DBSCAN)

	<i>Cluster 1</i> 34,11%	<i>Cluster 2</i> 9,78%	<i>Cluster 3</i> 17,18%	<i>Cluster 4</i> 37,75%
RECENCIA	116,84	223,21	276,08	353,65
FREQUENCIA	10,6	3	2	1
MONETARIO	214,73	52,19	34,77	17,15

3.4.2. Análises PPS

Com o intuito de proceder com a aplicação dos sistemas PPS, baseados na categorização dos produtos adquiridos pelos consumidores, optou-se por recorrer a estratégias mais indicadas para lidar com dados esparsos, tendo em atenção que 82,15% dos valores das variáveis patentes no Anexo A correspondem a zeros, devido a, como já referido, nenhum dos 17 mercados comerciais reunir compras de mais de 50% dos clientes. Mais concretamente, selecionaram-se os métodos de agrupamento *K-means* e *HAC (Hierarquical Agglomerative Clustering)*, aplicando a similaridade do cosseno como medida de distância, visto, segundo Dhillon e Modha (2001), esta constituir uma boa alternativa para trabalhar com a esparsidade e a alta dimensionalidade dos dados, sendo, assim, frequentemente utilizada nesse tipo de contextos. A métrica de distância referida, calculável através da regra visível na FÓRMULA 3.2, onde x_i e y_i correspondem a cada componente dos vetores em análise, com $i = 1, 2, \dots, n$ e n a igualar o número de dimensões, mede a similaridade de dois vetores como o cosseno do ângulo entre estes, ultrapassando, desta forma, a sensibilidade da distância euclidiana às normas dos mesmos (Xia et al., 2015).

$$\cos(\theta) = \frac{\sum_{i=1}^n x_i \times y_i}{\sqrt{\sum_{i=1}^n x_i^2} \times \sqrt{\sum_{i=1}^n y_i^2}} \quad (3.2)$$

3.4.2.1. K-means

Segundo Khan et al. (2021), o algoritmo *K-means* consiste num método de partição que divide o conjunto de dados em k grupos disjuntos, de forma a minimizar a soma dos quadrados. O método começa por seleccionar, de modo aleatório, k objetos, com cada um a representar, inicialmente, o centro de um *cluster*. De seguida, procede-se à associação dos restantes elementos ao seu centroide mais próximo e à redefinição dos centros dos *clusters* como a média de todos os pontos vinculados a cada agrupamento, com os procedimentos enunciados a serem repetidos até que não existam mudanças nos grupos formados em duas iterações consecutivas.

De modo a proceder com a execução da técnica *K-means* na presente dissertação, seguindo o estudo de Huang et al. (2020), que recorre ao método do cotovelo para a seleção do número ótimo de grupos, o total de *clusters* foi estipulado em 12, como indica o ponto de inflexão observável na FIGURA 3.19. Nos QUADROS 3.8 e 3.9 estão patentes as principais características dos *clusters* obtidos com a análise PPS efetuada através do algoritmo *K-means*, aplicado por meio da similaridade do cosseno, sendo possível verificar a existência de um agrupamento que regista compras em variadas categorias (*Cluster 9*) e onze segmentos que concentram, pelo menos, um terço dos seus gastos numa repartição comercial da farmácia.

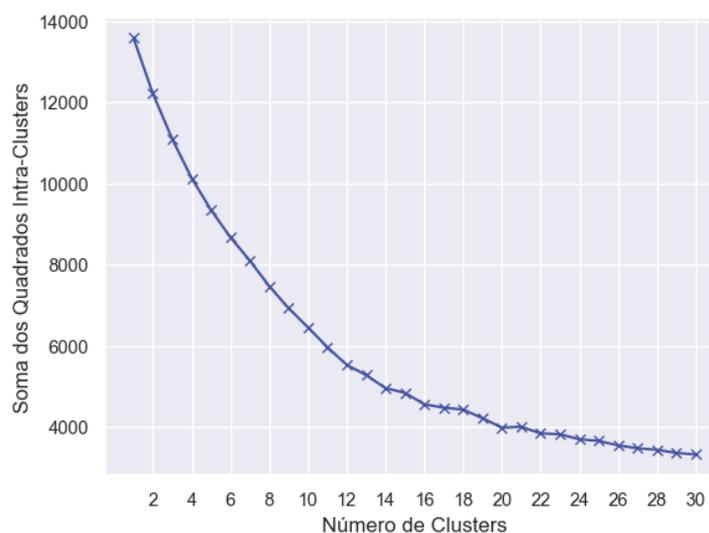


FIGURA 3.19. Método do cotovelo (*K-means*)

QUADRO 3.8. Média das variáveis de agrupamento por *cluster* (*K-means*) – *Clusters 1 a 6*

	Cluster 1 8,09%	Cluster 2 6,57%	Cluster 3 8,08%	Cluster 4 7,64%	Cluster 5 8,99%	Cluster 6 8,49%
Ajudas técnicas	0	0,44	0	0,03	0	0,02
Bebé e mamã	0	0	0	0,34	0	0

QUADRO 3.8. Continuação

	Cluster 1 8,09%	Cluster 2 6,57%	Cluster 3 8,08%	Cluster 4 7,64%	Cluster 5 8,99%	Cluster 6 8,49%
Contraceção e sexualidade	0	0	0	0	0	0
Cuidados específicos – olhos e ouvidos	0	0	0	0,02	<u>0,4</u>	0
Dermofarmácia, cosmética e acessórios	0,11	0,18	0,09	0,15	0,13	<u>0,93</u>
Higiene e cuidado oral	0	0,02	0	0,03	0,03	0
Homeopatia	0	0	0	0	0	0
Manipulação, soro fisiológico e soluções de irrigação	0	0	0	0	0	0
Medição de parâmetros e testes analíticos	0	0	0	0,03	0,01	0
Nutrição	0	0	0	0	0	0
Saúde animal	0	0	0	0	0	0
Sistema digestivo	0,05	0,05	<u>0,64</u>	0,05	0,06	0
Sistema nervoso e cessação tabágica	0,07	0,07	0,06	0,07	0,05	0,01
Sistema respiratório	0,11	0,11	0,1	0,11	0,12	0,05
Sistemas musculoesquelético e circulatório	<u>0,57</u>	0,05	0	0,04	0,06	0
Solares	0	0	0	0	0	0
Suplementos alimentares	0,08	0,07	0,1	0,12	0,12	0

QUADRO 3.9. Média das variáveis de agrupamento por cluster (K-means) – Clusters 7 a 12

	Cluster 7 7,64%	Cluster 8 8,76%	Cluster 9 12,15%	Cluster 10 8,46%	Cluster 11 7,79%	Cluster 12 7,35%
Ajudas técnicas	0	0,02	0,03	0	0	0
Bebé e mamã	0	0	0,03	0	0	0
Contraceção e sexualidade	0	0	0,03	0	0	0
Cuidados específicos – olhos e ouvidos	0	0,01	0,03	0	0	0
Dermofarmácia, cosmética e acessórios	0,1	0,1	0,18	0,08	0,03	0,15
Higiene e cuidado oral	0	0,02	0,04	0	0	<u>0,46</u>
Homeopatia	0	0	0,04	0	0	0
Manipulação, soro fisiológico e soluções de irrigação	0	0	0,03	0	0	0

QUADRO 3.9. Continuação

	Cluster 7 7,64%	Cluster 8 8,76%	Cluster 9 12,15%	Cluster 10 8,46%	Cluster 11 7,79%	Cluster 12 7,35%
Medição de parâmetros e testes analíticos	0	<u>0,48</u>	0,04	0	0	0
Nutrição	0	0	0,06	0	0	0
Saúde animal	0	0	0,01	0	0	0
Sistema digestivo	0	0,05	0,06	0	0	0,05
Sistema nervoso e cessação tabágica	0,05	0,07	0,06	<u>0,74</u>	0,01	0,07
Sistema respiratório	0,09	0,12	0,1	0,17	<u>0,96</u>	0,12
Sistemas musculo-esquelético e circulatório	0	0,05	0,05	0	0	0,05
Solares	0	0	0,07	0	0	0
Suplementos alimentares	<u>0,75</u>	0,07	0,14	0	0	0,09

3.4.2.2. HAC

Segundo Kazsoki e Hartmann (2020), o algoritmo HAC (*Hierarchical Agglomerative Clustering*) atua sob a premissa de que determinado objeto é mais similar a um ponto que lhe esteja mais próximo do que a um que lhe esteja mais afastado, utilizando uma abordagem *bottom-up*. A técnica começa por considerar cada elemento como um *cluster*, procedendo, de seguida, à junção dos dois grupos mais próximos, com este procedimento a ser repetido até que todos os objetos estejam incluídos num só agrupamento.

De modo a prosseguir com a execução da técnica HAC no presente estudo, definiu-se o método *average linkage*, que emite a distância média entre todos os pares de pontos de dois *clusters*, como a forma de cálculo da distância entre grupos, por ser, de acordo com Everitt et al. (2011), uma metodologia robusta, e, por isso, mais apropriada para lidar com a esparsidade e a alta dimensionalidade dos dados. Já seguindo o estudo de Sarparandeh e Hezarkhani (2017), que sugere a visualização do dendrograma para a seleção do número ótimo de segmentos, o total de *clusters* foi estipulado em 17, com o corte a ser definido em 0,5, como observável na FIGURA 3.20. Nos QUADROS 3.10, 3.11 e 3.12 estão patentes as principais características dos *clusters* obtidos com a análise PPS efetuada através do algoritmo HAC, aplicado por meio da similaridade do cosseno, sendo possível definir a existência de dezassete agrupamentos distinguidos por agregarem uma parte significativa das suas despesas numa determinada repartição comercial da farmácia.

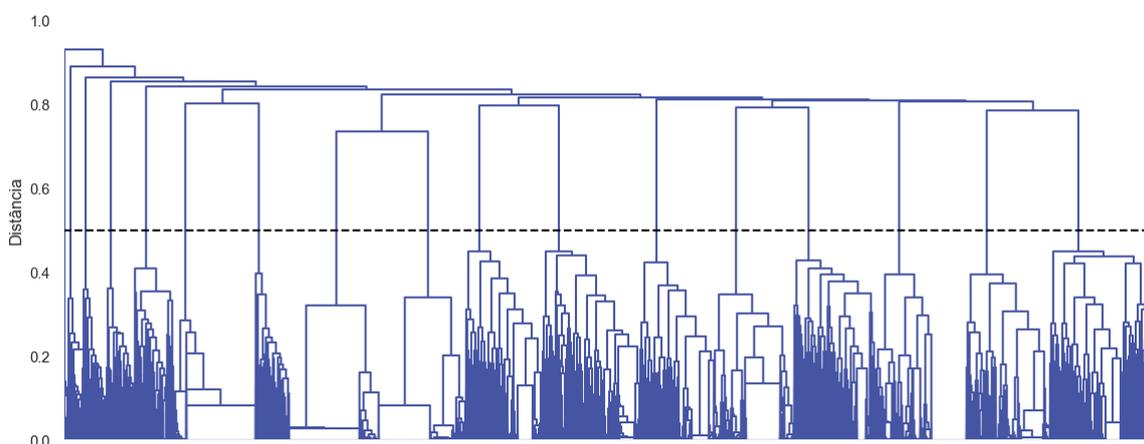


FIGURA 3.20. Dendrograma

QUADRO 3.10. Média das variáveis de agrupamento por *cluster* (HAC) – *Clusters* 1 a 6

	Cluster 1 9%	Cluster 2 9,7%	Cluster 3 6,98%	Cluster 4 7,65%	Cluster 5 6,73%	Cluster 6 8,34%
Ajudas técnicas	0,03	0,02	<u>0,4</u>	0	0	0
Bebé e mamã	0	0	0	0	0	0
Contraceção e sexualidade	0	0	0	0	0	0
Cuidados específicos – olhos e ouvidos	0,01	<u>0,37</u>	0	0	0	0
Dermofarmácia, cosmética e acessórios	0,12	0,13	0,16	0,11	0,14	0,11
Higiene e cuidado oral	0,03	0,03	0,03	0	<u>0,47</u>	0
Homeopatia	0	0	0	0	0	0
Manipulação, soro fisiológico e soluções de irrigação	0	0	0	0	0	0
Medição de parâmetros e testes analíticos	<u>0,45</u>	0,03	0	0	0	0
Nutrição	0	0	0	0	0	0
Saúde animal	0	0	0	0	0	0
Sistema digestivo	0,05	0,06	0,06	0,06	0,05	<u>0,61</u>
Sistema nervoso e cessação tabágica	0,07	0,05	0,08	0,07	0,07	0,06
Sistema respiratório	0,12	0,12	0,12	0,11	0,12	0,11
Sistemas musculoesquelético e circulatório	0,05	0,07	0,06	<u>0,57</u>	0,05	0
Solares	0	0	0	0	0	0
Suplementos alimentares	0,08	0,12	0,08	0,09	0,1	0,11

QUADRO 3.11. Média das variáveis de agrupamento por *cluster* (HAC) – *Clusters 7 a 12*

	Cluster 7 7,46%	Cluster 8 7,96%	Cluster 9 2,58%	Cluster 10 2,11%	Cluster 11 3,03%	Cluster 12 1,39%
Ajudas técnicas	0,03	0	0,06	0,02	0,02	0,02
Bebê e mamã	<u>0,33</u>	0	0,06	0,04	0,03	0,05
Contracepção e sexualidade	0	0	0	0	0	<u>0,26</u>
Cuidados específicos – olhos e ouvidos	0,02	0	0,03	0,03	0,03	0,03
Dermofarmácia, cosmética e acessórios	0,15	0,05	0,16	0,15	0,27	0,19
Higiene e cuidado oral	0,04	0	0,05	0,04	0,04	0,04
Homeopatia	0	0	0	<u>0,23</u>	0	0
Manipulação, soro fisiológico e soluções de irrigação	0	0	<u>0,17</u>	0	0	0
Medição de parâmetros e testes analíticos	0,03	0	0,03	0,03	0,03	0,03
Nutrição	0	0	0,01	0,01	0,01	0,01
Saúde animal	0	0	0	0	0	0
Sistema digestivo	0,05	0	0,06	0,05	0,04	0,05
Sistema nervoso e cessação tabágica	0,07	<u>0,73</u>	0,07	0,07	0,04	0,06
Sistema respiratório	0,11	0,22	0,12	0,12	0,07	0,1
Sistemas musculoesquelético e circulatório	0,04	0	0,06	0,06	0,05	0,05
Solares	0	0	0	0,01	<u>0,25</u>	0,01
Suplementos alimentares	0,12	0	0,12	0,14	0,12	0,1

QUADRO 3.12. Média das variáveis de agrupamento por *cluster* (HAC) – *Clusters 13 a 17*

	Cluster 13 7,58%	Cluster 14 3,47%	Cluster 15 7,33%	Cluster 16 0,45%	Cluster 17 8,25%
Ajudas técnicas	0	0,04	0	0,02	0
Bebê e mamã	0	0,03	0	0,01	0
Contracepção e sexualidade	0	0	0	0	0
Cuidados específicos – olhos e ouvidos	0	0,02	0	0,03	0
Dermofarmácia, cosmética e acessórios	<u>0,94</u>	0,14	0,11	0,15	0,12
Higiene e cuidado oral	0	0,04	0	0,01	0
Homeopatia	0	0	0	0,01	0

QUADRO 3.12. Continuação

	Cluster 13 7,58%	Cluster 14 3,47%	Cluster 15 7,33%	Cluster 16 0,45%	Cluster 17 8,25%
Manipulação, soro fisiológico e soluções de irrigação	0	0	0	0	0
Medição de parâmetros e testes analíticos	0	0,04	0	0,06	0
Nutrição	0	<u>0,2</u>	0	0	0
Saúde animal	0	0	0	<u>0,26</u>	0
Sistema digestivo	0	0,07	0	0,04	0
Sistema nervoso e cessação tabágica	0,06	0,07	0,06	0,06	0
Sistema respiratório	0	0,1	0,1	0,11	<u>0,88</u>
Sistemas musculoesquelético e circulatório	0	0,05	0	0,08	0
Solares	0	0	0	0,01	0
Suplementos alimentares	0	0,19	<u>0,73</u>	0,16	0

Resultados e discussão

A ponderação da relevância do presente estudo visa a apresentação dos resultados, com a apuração das métricas de avaliação de cada modelo e a posterior caracterização, para cada tipo de análise, do melhor agrupamento obtido, e a respetiva discussão, tendo em consideração a revisão de literatura realizada.

4.1. Resultados

A apreciação dos resultados dos modelos obtidos na atual dissertação efetuou-se com recurso a três métricas de avaliação, designadamente ao índice de Calinski-Harabasz, baseado no rácio das dispersões *inter-cluster* e *intra-cluster*, ao índice de Davies-Bouldin, estabelecido através da similaridade média dos segmentos com o respetivo agrupamento mais próximo, e ao coeficiente de silhueta, definido com recurso às diferenças entre as distâncias médias de determinado objeto aos elementos do próprio grupo e aos constituintes do agregado mais próximo (Gomes & Meisen, 2023; Salminen et al., 2023). Adicionalmente, para cada tipo de análise, procedeu-se à averiguação da afinidade das segmentações decorrentes dos diferentes modelos testados, utilizando, para tal, o índice de informação mútua ajustado e o índice de Rand ajustado, que avaliam, de modo respetivo, a concordância e a similaridade das mesmas (Ho et al., 2023).

4.1.1. Análises RFM

Atentando nas métricas de avaliação patentes no QUADRO 4.1, alusivas às análises RFM realizadas, alicerçadas na recência, na frequência e no valor monetário despendido pelos consumidores, verifica-se que a técnica CLARA apresenta uma *performance* consideravelmente superior ao algoritmo DBSCAN, dada a segmentação resultante do primeiro modelo evidenciar *clusters* com maior densidade e maior separação entre si, comprovável pelo índice de Calinski-Harabasz e pelo coeficiente de silhueta mais elevados e pelo índice de Davies-Bouldin mais reduzido. Constata-se, ainda, que os resultados dos dois métodos mencionados exibem uma similaridade moderada, com o índice de informação mútua ajustado e o índice de Rand ajustado a denotarem valores inferiores a 55%.

QUADRO 4.1. Métricas de avaliação dos resultados dos modelos CLARA e DBSCAN

	CLARA	DBSCAN
Índice de Calinski-Harabasz	15 059	9 026,4
Índice de Davies-Bouldin	0,94	2,1
Coefficiente de Silhueta	0,36	0,19
Índice de Informação Mútua Ajustado	0,52	
Índice de Rand Ajustado	0,54	

Considerando o melhor desempenho do algoritmo CLARA face ao método DBSCAN, optou-se por prosseguir com a caracterização da segmentação resultante do primeiro modelo referido, recorrendo, para tal, aos resultados observáveis nos QUADROS 3.6, 4.2 e 4.3. O *Cluster 1*, distinguido por apresentar um bom valor médio de recência (66 dias) e baixos valores médios de frequência (2 visitas) e de montante gasto (33,75 euros), agrupa 19,75% dos consumidores analisados, identificados como os “novos clientes”, por revelarem interesse no negócio, tendo efetuado compras recentemente, mas evidenciarem apenas uma pequena quantia despendida na farmácia, devido a serem fregueses há pouco tempo. Já o *Cluster 2*, tipificado por exibir um bom valor médio de recência (63 dias) e altos valores médios de frequência (13 visitas) e de quantia desembolsada (281,05 euros), alberga 25,22% dos compradores, classificados como os “clientes leais”, por exibirem um excelente envolvimento com o comércio e, conseqüentemente, representarem a maior fatia de lucro do comércio, de acordo com o princípio de Pareto. No que concerne ao *Cluster 3*, diferenciado por revelar um mau valor médio de recência (416 dias) e baixos valores médios de frequência (1 visita) e de montante despendido (15,4 euros), este engloba 33,24% dos fregueses estudados, denominados como os “clientes em risco”, por mostrarem sinais de abandono do negócio, sendo, no entanto, relevante tentar mantê-los, visto corresponderem a cerca de um terço dos utentes do estabelecimento. Relativamente ao *Cluster 4*, assinalado por denotar um mau valor médio de recência (356 dias) e valores médios de frequência (3 visitas) e de quantia despendida (63,71 euros) consideráveis, este abarca 21,79% dos consumidores, intitulados como os “clientes que necessitam de atenção”, por, apesar de manifestarem indícios de perda de interesse no comércio, não registando transações nos últimos tempos, revelarem um envolvimento significativo com o mesmo no passado, sendo, assim, bastante importante retê-los, dado o impacto expressivo que tiveram e podem voltar a ter na faturação da farmácia. Importa, ainda, salientar que os “clientes leais” e os “clientes em risco”, integrados nos *Clusters 2 e 3* na devida ordem, são, como expectável, os mais díspares entre si, com os primeiros a totalizarem apenas 36,47% de indivíduos do sexo masculino e 4,72% de utentes com residência fiscal fora do distrito de localização do estabelecimento comercial, face aos 44,58% e 24% registados, de modo respetivo, pelos segundos,

a mostrarem uma taxa média de lucro equivalente a somente 33,42%, uma quota média de remédios genéricos comprados igual a apenas 7,53% e uma percentagem média de medicamentos adquiridos correspondente a só 50,16%, perante os 37,13%, 8,39% e 61,4% assinalados, respetivamente, pelos últimos, e a albergarem clientes com uma média de idades igual a 57 anos, ante a média de 47 anos exibida pelo *Cluster 3*.

QUADRO 4.2. Frequência relativa das categorias das variáveis “CLIENTE_TIPO_PESSOA”, “CLIENTE_SEXO” e “RESIDENTE” por *cluster* (CLARA)

	CLIENTE_TIPO_PESSOA			CLIENTE_SEXO		RESIDENTE	
	Singular	Colectiva	Empregado	Feminino	Masculino	S	N
Cluster 1	99,91%	0,06%	0,03%	57,27%	42,73%	82,65%	17,35%
Cluster 2	99,79%	0,14%	0,07%	63,53%	36,47%	95,28%	4,72%
Cluster 3	99,84%	0,16%		55,42%	44,58%	76%	24%
Cluster 4	99,87%	0,13%		60,51%	39,49%	90,6%	9,4%

QUADRO 4.3. Média das variáveis “LUCRO_PERC”, “QUANTIDADE”, “GENERICOS_PERC”, “MEDICAMENTOS_PERC” e “IDADE” por *cluster* (CLARA)

	LUCRO_PERC	QUANTIDADE	GENERICOS_PERC	MEDICAMENTOS_PERC	IDADE
Cluster 1	35,79	3,15	7,5	58,75	46,6
Cluster 2	33,42	22,37	7,53	50,16	56,67
Cluster 3	37,13	1,58	8,39	61,4	46,82
Cluster 4	35,47	5,75	7,27	54,51	51,25

4.1.2. Análises PPS

Observando as métricas de avaliação patentes no QUADRO 4.4, alusivas às análises PPS concretizadas, baseadas na categorização dos produtos adquiridos pelos consumidores, conclui-se que, embora o valor mais reduzido do índice de Davies-Bouldin aponte para a escolha do algoritmo HAC, a técnica *K-means* exhibe, no geral, uma *performance* superior ao mesmo, visto a respetiva segmentação resultante evidenciar *clusters* com maior densidade e maior separação entre si, de acordo com os valores mais elevados do índice de Calinski-Harabasz e do coeficiente de silhueta. Afere-se, também, que os resultados dos dois métodos referidos revelam uma similaridade elevada, com o índice de informação mútua ajustado e o índice de Rand ajustado a revelarem valores superiores a 76%.

QUADRO 4.4. Métricas de avaliação dos resultados dos modelos K-means e HAC

	K-means	HAC
Índice de Calinski-Harabasz	2 284,72	1 935,33
Índice de Davies-Bouldin	1,65	1,13
Coefficiente de Silhueta	0,31	0,29
Índice de Informação Mútua Ajustado	0,82	
Índice de Rand Ajustado	0,76	

Atendendo ao melhor desempenho do algoritmo K-means perante a técnica HAC, optou-se por proceder com a caracterização da segmentação resultante do primeiro modelo mencionado, utilizando, para tal, os resultados disponíveis nos QUADROS 3.8, 3.9, 4.5 e 4.6. O método em causa identifica doze grupos de clientes, designadamente um agrupamento que regista compras em variadas categorias e onze segmentos que concentram, pelo menos, um terço dos seus gastos numa única repartição comercial da farmácia. Deste modo, o *Cluster 9* alberga 12,15% dos fregueses analisados, caracterizados por adquirirem produtos de todos os setores comerciais do estabelecimento, centralizando, no mínimo, 1% dos seus consumos em cada um dos mercados, em termos médios, ao passo que o *Cluster 1* inclui 8,09% dos compradores observados, distinguidos por acumularem, em média, 57% dos seus gastos na esfera “Sistemas musculoesquelético e circulatório”, o *Cluster 2* abarca 6,57% dos clientes, identificados por aglomerarem, em média, 44% das suas despesas na divisão “Ajudas técnicas”, o *Cluster 3* engloba 8,08% dos utentes, reconhecidos por congregarem, em média, 64% das suas expensas no departamento “Sistema digestivo”, o *Cluster 4* compreende 7,64% dos consumidores, diferenciados por reterem, em média, 34% dos seus dispêndios no campo “Bebé e mamã”, o *Cluster 5* contém 8,99% dos fregueses, discernidos por juntarem, em média, 40% dos seus gastos no domínio “Cuidados específicos – olhos e ouvidos”, o *Cluster 6* acolhe 8,49% dos compradores, assinalados por acumularem, em média, 93% das suas despesas na categoria “Dermofarmácia, cosmética e acessórios”, o *Cluster 7* acomoda 7,64% dos clientes, denotados por aglomerarem, em média, 75% das suas expensas no espaço “Suplementos alimentares”, o *Cluster 8* integra 8,76% dos utentes, caracterizados por congregarem, em média, 48% dos seus dispêndios no ramo “Medição de parâmetros e testes analíticos”, o *Cluster 10* alberga 8,46% dos consumidores, distinguidos por reterem, em média, 74% dos seus gastos na repartição “Sistema nervoso e cessação tabágica”, o *Cluster 11* inclui 7,79% dos fregueses, identificados por juntarem, em média, 96% das suas despesas no âmbito “Sistema respiratório”, e o *Cluster 12* abarca 7,35% dos compradores, reconhecidos por acumularem, em média, 46% das suas expensas no segmento “Higiene e cuidado oral”. Importa referir que o *Cluster 11*, que alberga os principais fregueses da categoria “Sistema

respiratório”, é aquele que possui, de certo modo, as características mais distintivas, totalizando as maiores presenças de indivíduos do sexo masculino (48,44%) e de utentes com residência fiscal fora do distrito de localização do estabelecimento comercial (22,9%), exibindo as maiores percentagens médias de lucro (41,06%) e de medicamentos adquiridos (93,35%), a par do *Cluster 10*, que inclui os maiores consumidores da secção “Sistema nervoso e cessação tabágica”, e demonstrando, juntamente ao *Cluster 6*, que acolhe os melhores clientes da esfera “Dermofarmácia, cosmética e acessórios”, o menor número médio de produtos comprados (2 unidades) e a maior quota média de remédios genéricos adquiridos (10,61%). Em contrapartida, constata-se que o *Cluster 9*, que inclui os clientes que adquirem produtos de todos os setores, denota a menor presença de sujeitos do género masculino (30,36%) e a maior quantidade média de produtos adquiridos (20 unidades), o *Cluster 8*, que integra os maiores fregueses do espaço “Medição de parâmetros e testes analíticos”, apresenta a menor taxa média de lucro (27,72%) e o *Cluster 7*, que acomoda os principais consumidores do segmento “Suplementos alimentares”, revela a menor percentagem média de medicamentos adquiridos (31,56%). Adicionalmente, verifica-se que os *Clusters 1* e *5*, que englobam, na devida ordem, os melhores compradores das repartições “Sistemas musculo-esquelético e circulatório” e “Cuidados específicos – olhos e ouvidos”, registam a maior média de idades, correspondente a 58 e a 59 anos, respetivamente, enquanto o *Cluster 4*, que abarca os maiores utentes do departamento “Bebé e mamã”, assinala a menor média, equivalente a 38 anos.

QUADRO 4.5. Frequência relativa das categorias das variáveis “CLIENTE_TIPO_PESSOA”, “CLIENTE_SEXO” e “RESIDENTE” por *cluster* (K-means)

	CLIENTE_TIPO_PESSOA			CLIENTE_SEXO		RESIDENTE	
	Singular	Colectiva	Empregado	Feminino	Masculino	S	N
Cluster 1	99,64%	0,36%		58,52%	41,48%	89,23%	10,77%
Cluster 2	99,91%	0,09%		54,08%	45,92%	86,51%	13,49%
Cluster 3	100%			54,74%	45,26%	83,95%	16,05%
Cluster 4	99,47%	0,45%	0,08%	63,35%	36,65%	88,28%	11,72%
Cluster 5	99,87%	0,13%		57,3%	42,7%	89,76%	10,24%
Cluster 6	100%			62,3%	37,7%	79,68%	20,32%
Cluster 7	100%			61,36%	38,64%	83,42%	16,58%
Cluster 8	99,87%	0,13%		52,29%	47,71%	89,15%	10,85%
Cluster 9	99,71%	0,19%	0,1%	69,84%	30,16%	88,72%	11,28%
Cluster 10	100%			54,46%	45,54%	79,94%	20,06%
Cluster 11	99,93%		0,07%	51,56%	48,44%	77,1%	22,9%
Cluster 12	99,84%	0,16%		61,62%	38,38%	86,95%	13,05%

QUADRO 4.6. Média das variáveis “LUCRO_PERC”, “QUANTIDADE”, “GENERICOS_PERC”, “MEDICAMENTOS_PERC” e “IDADE” por cluster (K-means)

	LUCRO_PERC	QUANTIDADE	GENERICOS_PERC	MEDICAMENTOS_PERC	IDADE
Cluster 1	37,39	6,75	6,62	81,6	57,69
Cluster 2	35,08	8,56	6,62	37,01	52,08
Cluster 3	35,69	3,68	5,99	86,91	51,32
Cluster 4	32,97	11,86	6,32	36,72	38,47
Cluster 5	35,06	10,09	8	45,58	59,15
Cluster 6	37,19	2,39	10,48	52,6	46,08
Cluster 7	34,91	4,16	7,39	31,56	47,13
Cluster 8	27,72	11,2	7,81	35,26	55,97
Cluster 9	35,11	19,56	7,31	40,97	53,66
Cluster 10	40,13	2,74	8,18	96,24	42,88
Cluster 11	41,06	1,82	10,61	93,35	45,89
Cluster 12	35	7,79	7,03	43,96	49,71

4.2. Discussão

A presente dissertação concretiza, através de uma análise RFM, assente na recência, na frequência e no valor monetário despendido, e de um modelo PPS, baseado na categorização dos produtos adquiridos, a divisão do mercado de uma farmácia comunitária, visando, deste modo, colmatar a escassez de estudos relativos à segmentação comportamental de clientes, efetuada a partir do historial de compras dos mesmos, no contexto referido, com a literatura científica existente a circunscrever-se, através de questionários aplicados aos compradores, à perspetiva psicográfica e à abordagem referente ao valor percecionado para efetuar segmentações de mercado no setor de farmácia. Os resultados demonstram, assim, a partir da identificação de quatro grupos de fregueses com diferentes graus de envolvimento com o negócio – “novos clientes”, “clientes leais”, “clientes em risco” e “clientes que necessitam de atenção” –, obtidos com recurso ao sistema RFM, e de doze conjuntos de compradores com vontades distintas – um agrupamento geral e onze segmentos especializados –, alcançados através do método PPS, que a segmentação comportamental, baseada nas análises em causa, é aplicável à conjuntura de farmácia comunitária, permitindo adquirir *insights* valiosos que possibilitam o aumento do sucesso das estratégias de *marketing*, ao viabilizarem categorizações dos consumidores de acordo com a sua relevância para o comércio e com as suas necessidades.

No que respeita à aplicação da análise RFM no atual estudo, a mesma atendeu ao trabalho de Doğan et al. (2018), onde o método é empregado a uma cadeia de lojas de retalho de desporto turca, de modo a redefinir, através dos algoritmos de *clustering two-step* e *K-means*, os estatutos de cliente de bronze, de ouro e *premium*, previamente estabelecidos apenas de acordo com o fator monetário. As principais diferenças entre os ensaios decorrem do facto de, no presente projeto, de forma a preservar a granularidade e a precisão dos dados, terem sido utilizados os valores reais da recência, da frequência e da quantia monetária, ao invés de ter sido efetuada a sua catalogação em níveis distintos, onde passam a assumir a estrutura de variáveis categóricas. Assim, a alteração referida obrigou, ainda, ao uso de algoritmos mais apropriados para lidar com dados com alta dispersão, pelo que se optou por recorrer à técnica CLARA, por ser mais robusta a casos extremos (Kaufman & Rousseeuw, 1990), e ao método DBSCAN, por permitir a deteção dos mesmos.

Alusivamente à execução do modelo PPS, a mesma concretizou-se mediante o artigo de Sokol e Holý (2021), onde o sistema é aplicado, a partir do algoritmo *K-means*, a uma cadeia de drogarias checa, com 55 tipologias existentes de produtos, sendo que, no presente estudo, vista a esparsidade e alta dimensionalidade dos dados, se escolheu recorrer aos métodos de agrupamento *K-means* e HAC usando a similaridade do cosseno como medida de distância, por esta constituir uma boa alternativa para trabalhar com as referidas características da informação (Dhillon & Modha, 2001). À semelhança do ensaio mencionado, a atual dissertação deteta um único agrupamento geral, que adquire artigos pertencentes a diversas categorias, e os restantes grupos especializados, com compras que consistem maioritariamente num género específico de bens, o que demonstra que, independentemente dos contextos completamente distintos analisados, o comportamento dos clientes se assemelha nesse sentido.

CAPÍTULO 5

Conclusões

O presente estudo visa a segmentação do mercado de uma farmácia comunitária portuguesa, localizada no distrito de Santarém, utilizando, deste modo, os algoritmos CLARA e DBSCAN para concretizar as análises RFM, assentes na recência, na frequência e no valor monetário despendido, e as técnicas *K-means* e HAC para executar os modelos PPS, baseados na categorização dos produtos adquiridos. Os resultados do melhor sistema RFM obtido, decorrente do método CLARA, apontam para a existência de quatro grupos de clientes com diferentes graus de envolvimento com o negócio, nomeadamente os “novos clientes”, os “clientes leais”, os “clientes em risco” e os “clientes que necessitam de atenção”. Já os produtos da melhor análise PPS conseguida, resultante do algoritmo *K-means*, sugerem a presença de doze conjuntos de compradores, designadamente de um agrupamento que adquire artigos pertencentes a diversas categorias e onze agregados que concentram, pelo menos, um terço dos seus gastos numa única secção do estabelecimento (“Sistemas musculoesquelético e circulatório”, “Ajudas técnicas”, “Sistema digestivo”, “Bebé e mamã”, “Cuidados específicos – olhos e ouvidos”, “Dermofarmácia, cosmética e acessórios”, “Suplementos alimentares”, “Medição de parâmetros e testes analíticos”, “Sistema nervoso e cessação tabágica”, “Sistema respiratório” ou “Higiene e cuidado oral”). Assim sendo, a atual dissertação não só permite constatar que a segmentação comportamental, baseada nos modelos mencionados, é aplicável à conjuntura de farmácia comunitária, colmatando a escassez de literatura relativa à abordagem comportamental, efetuada a partir do historial de compras, para divisão de mercado no contexto referido, como também possibilita a categorização dos consumidores de acordo com a sua relevância para o comércio e com as suas necessidades, revelando, através do sistema RFM, o interesse dos “clientes leais” e dos “clientes que necessitam de atenção” para o negócio, por, de modo respetivo, representarem a maior fonte de lucro do estabelecimento e demonstrarem um impacto expressivo na faturação da farmácia no passado, que, com as devidas estratégias de retenção, se pode voltar a repetir no futuro, e, através da análise PPS, o valor do segmento que adquire artigos pertencentes a diversas categorias e dos agrupamentos que albergam os principais fregueses das esferas “Bebé e mamã”, “Cuidados específicos – olhos e ouvidos” e “Medição de parâmetros e testes analíticos”, por, em termos médios, cada cliente dos mesmos, adquirir mais de 10 produtos no comércio.

A importância do estudo prende-se com o facto da divisão de clientes ser um procedimento fulcral para o sucesso das estratégias de *marketing*, ao facilitar a compreensão das necessidades dos compradores e, conseqüentemente, ao auxiliar os negócios a atender às demandas do mercado e a destacar-se da sua concorrência. Assim, a presente investigação, ao atestar que a segmentação comportamental é aplicável à conjuntura de farmácia comunitária, possibilita, ao gestor do estabelecimento, definir táticas para atrair o público-alvo, detetar nichos de mercado, otimizar os resultados das campanhas publicitárias e os recursos financeiros utilizados nas mesmas, desenvolver ofertas personalizadas para os diferentes segmentos encontrados e aumentar a lealdade do consumidor e os lucros conseguidos.

No que concerne às limitações da atual dissertação, denota-se que as estas se detêm com o facto do agrupamento dos clientes, efetuado segundo o seu grau de envolvimento com o negócio e as suas necessidades, se executar apenas mediante os produtos não sujeitos a receita médica adquiridos, pelo que, em termos exemplificativos, um consumidor que compra 25 medicamentos que requerem prescrição médica e apenas 2 itens não sujeitos à mesma pode ser categorizado no grupo dos “clientes em risco”, ao invés de na esfera dos “clientes leais”.

Por último, sugere-se, como linha de investigação futura, a execução de análises RFM, assentes na recência, na frequência e no valor monetário despendido, e de modelos PPS, baseados na categorização dos produtos adquiridos, a outras farmácias comunitárias, de forma a aferir se existe um comportamento-tipo do mercado deste setor, isto é, se, tendencialmente, a clientela de diferentes estabelecimentos exhibe hábitos semelhantes ou não. Adicionalmente, propõe-se, também, o desenvolvimento de uma ferramenta que possibilite o acesso automático a este tipo de resultados, de forma a permitir a associação de cada comprador ao respetivo grupo para a aplicação de campanhas publicitárias de farmácia mais dirigidas.

Referências Bibliográficas

- Ansoff, H. I. (1957). Strategies for Diversification. *Harvard Business Review*, 113-124.
- Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering*, 60(1), 208-221. <https://doi.org/10.1016/j.datak.2006.01.013>
- Carroll, N. V., & Gagon, J. P. (1983). Identifying Consumer Segments in Health Services Markets: An Application of Conjoint and Cluster Analyses to the Ambulatory Care Pharmacy Market. *Journal of Health Care Marketing*, 3(3), 22-34.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. SPSS.
- Chereches, M. C., Popa, C., & Buicu, C. F. (2021). Socio-Demographic Profile of the Romanian Patient Regarding the Factors Influencing the Choice of a Pharmacy Using Logistic Regression and Factor Analysis. *Farmacia*, 69(6), 1182-1188. <https://doi.org/10.31925/farmacia.2021.6.24>
- Cooil, B., Aksoy, L., & Keiningham, T. L. (2008). Approaches to Customer Segmentation. *Journal of Relationship Marketing*, 6(3-4), 9-39. https://doi.org/10.1300/J366v06n03_02
- Dhillon, I. S., & Modha, D. S. (2001). Concept Decompositions for Large Sparse Text Data Using Clustering. *Machine Learning*, 42, 143-175. <https://doi.org/10.1023/A:1007612920971>
- Doğan, O., Atçin, E., & Bulut, Z. A. (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. <http://www.ijceas.com/index.php/ijceas/article/view/174>
- Enesi, I., Liço, L., Biberaj, A., & Shahu, D. (2021). Analysing Clustering Algorithms Performance in CRM Systems. Em International Conference on Enterprise Information Systems (Org.). *Proceedings of the 23rd International Conference on Enterprise Information Systems (ICEIS 2021)*, 1, 803-809. 23rd International Conference on Enterprise Information Systems. <https://doi.org/10.5220/0010511008030809>
- Ernawati, E., Baharin, S. S. K., & Kasmin, F. (2020). A review of data mining methods in RFM-based customer segmentation. Em Universitas Kanjuruhan Malang (Org.). *Journal of Physics: Conference Series*, 1869. 2nd Annual Conference of Science and Technology (ANCOSET 2020), Malang, Indonesia. <https://doi.org/10.1088/1742-6596/1869/1/012085>
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Em Association for the Advancement of Artificial Intelligence (Org.). *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226-231. KDD-96: The Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, Estados Unidos da América.

- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Hierarchical clustering. In D. J. Balding, N. A. C. Cressie, G. M. Fitzmaurice, H. Goldstein, G. Molenberghs, D. W. Scott, ... S. Weisberg (Eds.), *Cluster Analysis* (pp. 71-110). Londres, Reino Unido: John Wiley & Sons, Inc.
- Gomes, M. A., & Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Information Systems and e-Business Management*, 21, 527-570. <https://doi.org/10.1007/s10257-023-00640-4>
- Gurău, C. (2005). Pharmaceutical marketing on the internet: marketing techniques and customer profile. *Journal of Consumer Marketing*, 22(7), 421-428. <https://doi.org/10.1108/07363760510631165>
- Hiziroglu, A. (2013). Soft computing applications in customer segmentation: State-of-art review and critique. *Expert Systems with Applications*, 40(16), 6491-6507. <https://doi.org/10.1016/j.eswa.2013.05.052>
- Ho, T., Nguyen, S., Nguyen, H., Nguyen, N., Man, D., & Le, T. (2023). An Extended RFM Model for Customer Behaviour and Demographic Analysis in Retail Industry. *Business Systems Research*, 14(1), 26-53. <https://doi.org/10.2478/bsrj-2023-0002>
- Huang, Y., Zhang, M., & He, Y. (2020). Research on improved RFM customer segmentation model based on K-Means algorithm. Em International Conference on Computational Intelligence and Applications (Org.). *2020 5th International Conference on Computational Intelligence and Applications (ICCIA)*, 24-27. 2020 5th International Conference on Computational Intelligence and Applications (ICCIA 2020). <https://doi.org/10.1109/ICCIA49625.2020.00012>
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, Inc.
- Kazsoki, A. S., & Hartmann, B. (2020). Hierarchical Agglomerative Clustering of Selected Hungarian Medium Voltage Distribution Networks. *Acta Polytechnica Hungarica*, 17(4), 201-219. <https://doi.org/10.12700/APH.17.4.2020.4.11>
- Kevrekidis, D. P., Minarikova, D., Markos, A., Malovecka, I., & Minarik, P. (2018). Community pharmacy customer segmentation based on factors influencing their selection of pharmacy and over-the-counter medicines. *Saudi Pharmaceutical Journal*, 26(1), 33-43. <https://doi.org/10.1016/j.jsps.2017.11.002>
- Khan, R. H., Dofadar, D. F., & Alam, M. G. R. (2021). Explainable Customer Segmentation Using K-means Clustering. Em Institute of Electrical and Electronics Engineers (Org.). *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 639-643. IEEE UEMCON 2021, Nova Iorque, Estados Unidos da América. <https://doi.org/10.1109/UEMCON53757.2021.9666609>
- Koli, S., Singh, R., Mishra, R., & Badhani, P. (2023). Imperative role of customer segmentation technique for customer retention using machine learning techniques. Em G.L. Bajaj Institute of Technology & Management (Org.). *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, 243-248. 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), Greater Noida, Índia. <https://doi.org/10.1109/AISC56616.2023.10085487>
- Kotler, P., & Armstrong, G. (1999). *Principles of Marketing* (Vol. 1). Prentice Hall.

- Liao, J., Jantan, A., Ruan, Y., & Zhou, C. (2022). Multi-Behavior RFM Model Based on Improved SOM Neural Network Algorithm for Customer Segmentation. *IEEE Access*, *10*, 122501-122512. <https://doi.org/10.1109/ACCESS.2022.3223361>
- Lostakova, H., Pecinova, Z., Hraska, V., Crhakova, K. (2015). Customer Value Attributes in Purchases of Otc Drugs and Food Supplements in Brick-And-Mortar Pharmacies in the Czech Republic. Em SGEM Social (Org.). *SGEM 2015 International Multidisciplinary Scientific Conference on Social Sciences and Arts*, 667–674. 2nd International Multidisciplinary Scientific Conference on Social Sciences and Arts, Albena, Bulgária. <https://doi.org/10.5593/sgemsocial2015/b22/s7.086>
- Mahfuza, R., Islam, N., Toyeb, M., Emon, M. A. F., Chowdhury, S. A., & Alam, M. G. R. (2022). LRFMV: An efficient customer segmentation model for superstores. *PLoS ONE*, *17*(12). <https://doi.org/10.1371/journal.pone.0279262>
- McDonald, M., & Dunbar, I. (2004). *Market Segmentation: How to do it, how to profit from it*. Butterworth-Heinemann.
- Nakano, S., & Kondo, F. N. (2018). Customer segmentation with purchase channels and media touchpoints using single source panel data. *Journal of Retailing and Consumer Services*, *41*, 142-152. <https://doi.org/10.1016/j.jretconser.2017.11.012>
- Nguyen, S. P. (2021). Deep customer segmentation with applications to a Vietnamese supermarkets' data. *Soft Computing*, *25*, 7785-7793. <https://doi.org/10.1007/s00500-021-05796-0>
- Patak, M., Lostakova, H., Curdova, M., & Vlckova, V. (2014). The E-Pharmacy Customer Segmentation Based on the Perceived Importance of the Retention Support Tools. *Procedia – Social and Behavioral Sciences*, *150*, 552-562. <https://doi.org/10.1016/j.sbspro.2014.09.075>
- Pribadi, P., Pangestuti, R. C. A., & Daniswara, N. J. (2020). Preference of pharmacies selection and OTC medicine purchase with customer segmentation approach: A case study in magelang. *International Journal of Scientific and Technology Research*, *9*(2), 2875-2879.
- Qian, C., Yang, M., Li, P., & Li, S. (2018). Application of Customer Segmentation for Electronic Toll Collection: A Case Study. *Journal of Advanced Transportation*, *2018*, 1-9. <https://doi.org/10.1155/2018/3635107>
- Salminen, J., Mustak, M., Sufyan, M., & Jansen, B. J. (2023). How can algorithms help in segmenting users and customers? A systematic review and research agenda for algorithmic customer segmentation. *Journal of Marketing Analytics*, *11*, 677-692. <https://doi.org/10.1057/s41270-023-00235-5>
- Sarparandeh, M., & Hezarkhani, A. (2017). Application of unsupervised pattern recognition approaches for exploration of rare earth elements in Se-Chahun iron ore, central Iran. *Geoscientific Instrumentation Methods and Data Systems*, *6*(2), 537-546. <https://doi.org/10.5194/gi-6-537-2017>
- Sayan, I. U., Demirdag, M., Yuceturk, G., & Yalcinkaya, S. M. (2022). A Review of Customer Segmentation Methods: The Case of Investment Sector. Em Institute of Electrical and Electronics Engineers (Org.). *2022 IEEE 5th International Conference on Big Data and Artificial Intelligence (BDAI)*, 200-204. The 5th International Conference on Big Data and Artificial Intelligence, Fuzhou, China. <https://doi.org/10.1109/BDAI56143.2022.9862801>

- Schommer, J. C., & Gaither, C. A. (2014). A segmentation analysis for pharmacists' and patients' views of pharmacists' roles. *Research in Social and Administrative Pharmacy*, 10(3), 508-528. <http://dx.doi.org/10.1016/j.sapharm.2013.10.004>
- Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. *ACM Transactions on Database Systems*, 42(3), 1-21. <https://doi.org/10.1145/3068335>
- Sista, R., Singh, R., Kumawat, S. K., & Dhanare, R. (2021). Techniques used by E-commerce industries for Customer analysis. Em International Conference on Computer Communication and Informatics (Org.). *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 1-6. 2021 International Conference on Computer Communication and Informatics (ICCCI -2021), Coimbatore, Índia. <https://doi.org/10.1109/ICCCI50826.2021.9402270>
- Smith, W. R. (1956). Product Differentiation and Market Segmentation as Alternative Marketing Strategies. *Journal of Marketing*, 21(1), 3-8. <https://doi.org/10.2307/1247695>
- Sokol, O., & Holý, V. (2021). The role of shopping mission in retail customer segmentation. *International Journal of Market Research*, 63(4), 454-470. <https://doi.org/10.1177/1470785320921011>
- Statista. (2024, janeiro). *Revenue of the worldwide pharmaceutical market from 2001 to 2023*. <https://www.statista.com/statistics/263102/pharmaceutical-market-worldwide-revenue-since-2001/>
- Weilbaker, D. C., Crocker, K. E., & Clabaugh, M. G. (1989). Retail Pharmacy Patronage: Mom and Pop Vs The Chain. Em Academy of Marketing Science (Org.). *Proceedings of the 1989 Academy of Marketing Science (AMS) Annual Conference*, 588-591. 1989 AMS Annual Conference, Orlando, Flórida, Estados Unidos da América. https://doi.org/10.1007/978-3-319-17055-8_120
- Xia, P., Zhang, L., & Li, F. (2015). Learning similarity with cosine similarity ensemble. *Information Sciences*, 307, 39-52. <https://doi.org/10.1016/j.ins.2015.02.024>
- Zakrzewska, D., & Murlewski, J. (2005). Clustering Algorithms for Bank Customer Segmentation. Em International Conference on Intelligent Systems Design and Applications (Org.). *Proceedings of the 2005 5th International Conference on Intelligent Systems Design and Applications (ISDA'05)*, 197-202. 5th International Conference on Intelligent Systems Design and Applications (ISDA 2005), Varsóvia, Polónia. <https://doi.org/10.1109/ISDA.2005.33>
- Zhou, J., Wei, J., & Xu, B. (2021). Customer segmentation by web content mining. *Journal of Retailing and Consumer Services*, 61. <https://doi.org/10.1016/j.jretconser.2021.102588>
- Zhou, X., Zhang, Z., & Lu, Y. (2011). Review of Customer Segmentation method in CRM. Em The University of Western Australia (Org.). *2011 International Conference on Computer Science and Service System (CSSS)*, 4033-4035. 2011 International Conference on Computer Science and Service System (CSSS), Nanjing, China. <https://doi.org/10.1109/CSSS.2011.5974617>

Anexo A

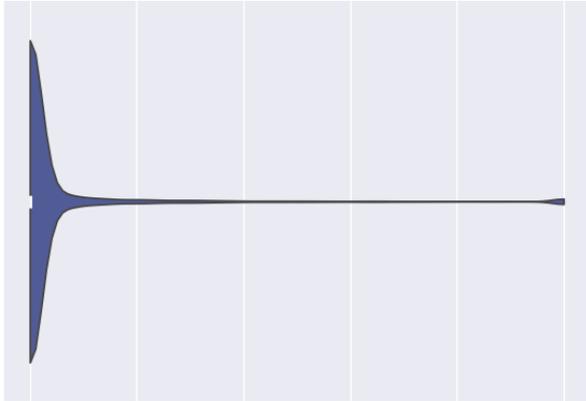


FIGURA A.21. Distribuição da variável "Ajudas técnicas"

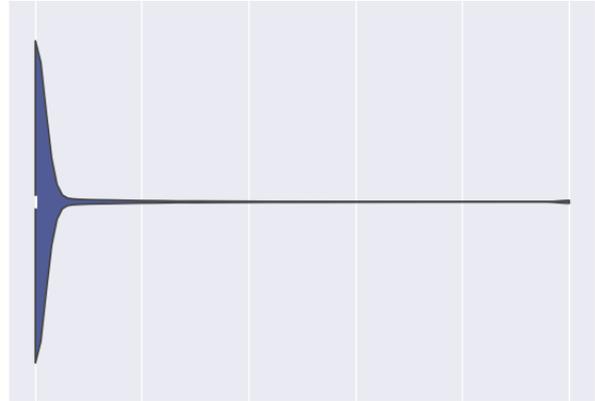


FIGURA A.22. Distribuição da variável "Bebé e mamã"

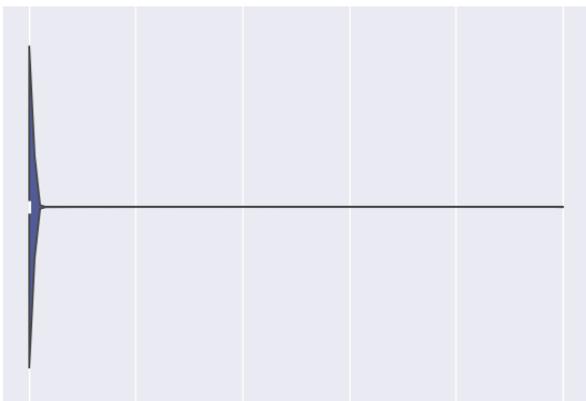


FIGURA A.23. Distribuição da variável "Contraceção e sexualidade"

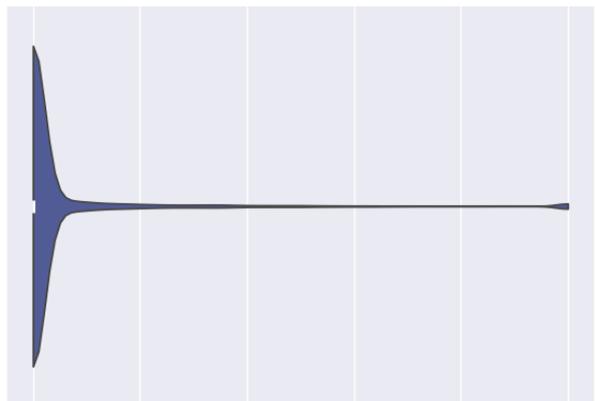


FIGURA A.24. Distribuição da variável "Cuidados específicos – olhos e ouvidos"

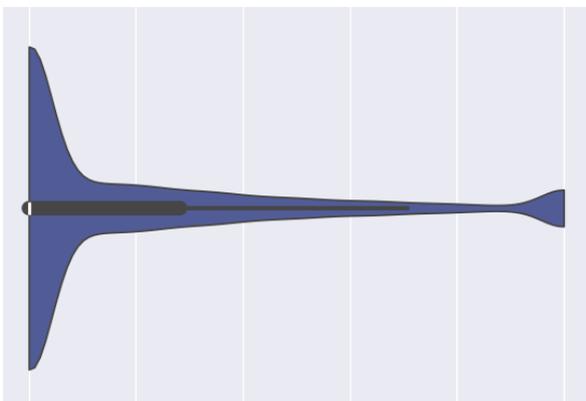


FIGURA A.25. Distribuição da variável "Dermofarmácia, cosmética e acessórios"

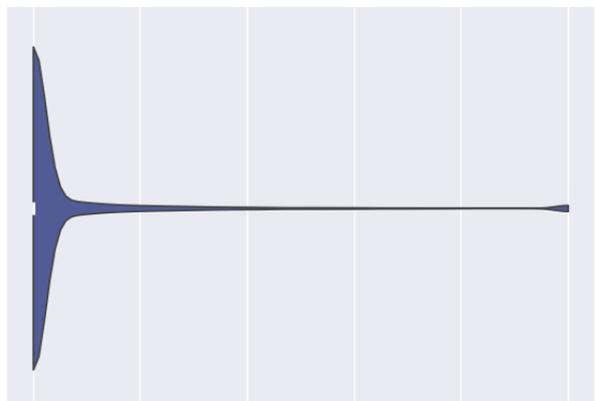


FIGURA A.26. Distribuição da variável "Higiene e cuidado oral"

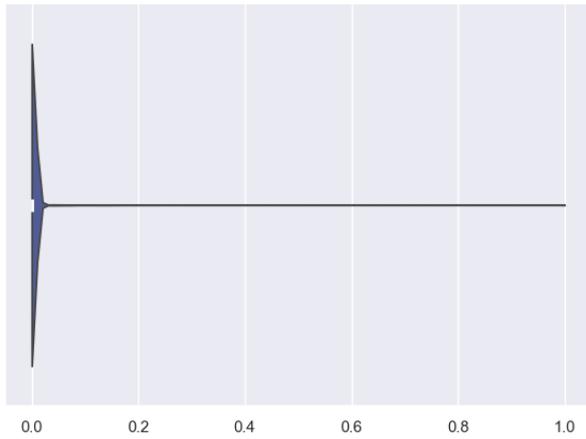


FIGURA A.27. Distribuição da variável "Homeopatia"

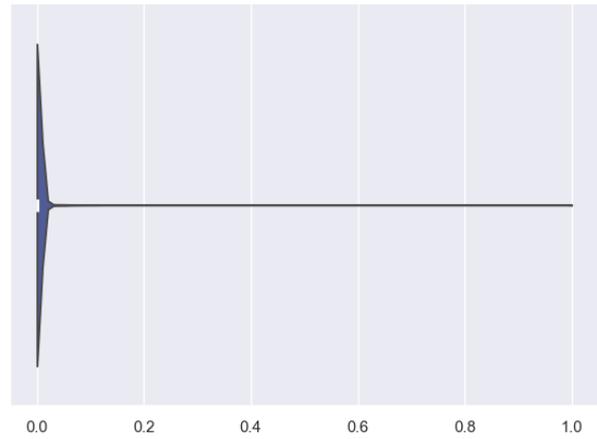


FIGURA A.28. Distribuição da variável "Manipulação, soro fisiológico e soluções de irrigação"

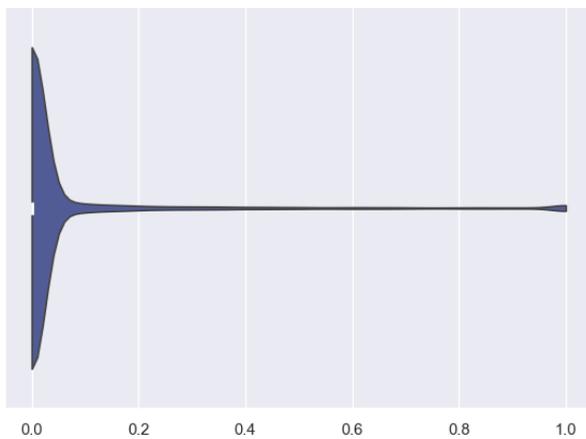


FIGURA A.29. Distribuição da variável "Medição de parâmetros e testes analíticos"

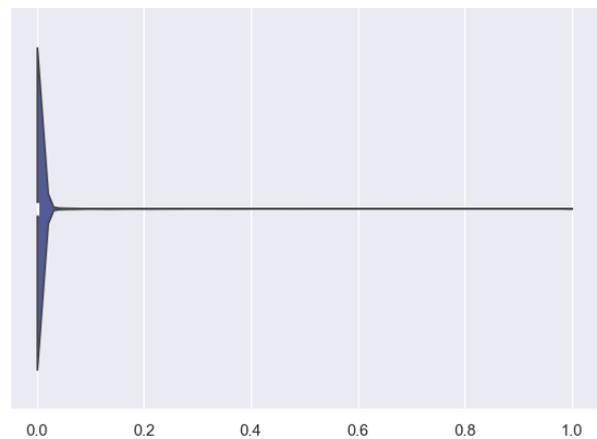


FIGURA A.30. Distribuição da variável "Nutrição"

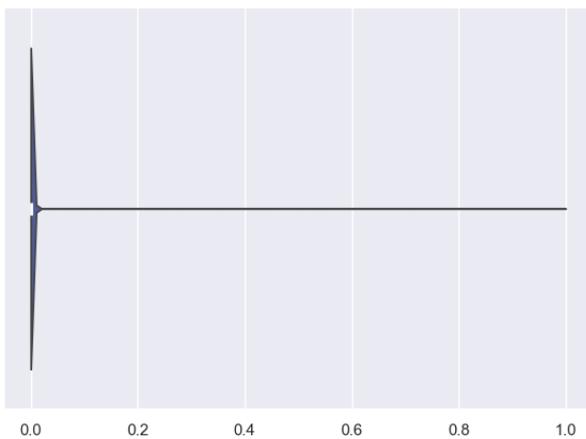


FIGURA A.31. Distribuição da variável "Saúde animal"

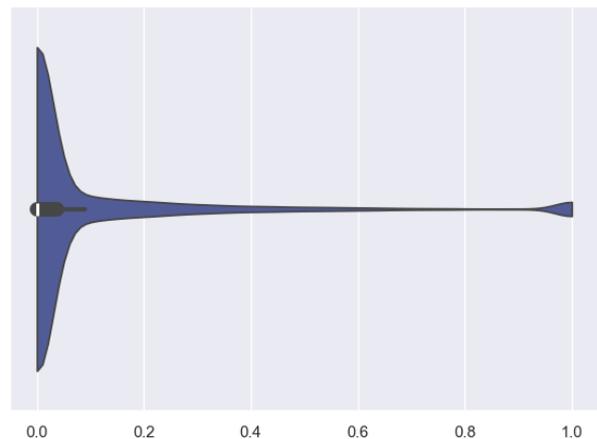


FIGURA A.32. Distribuição da variável "Sistema digestivo"

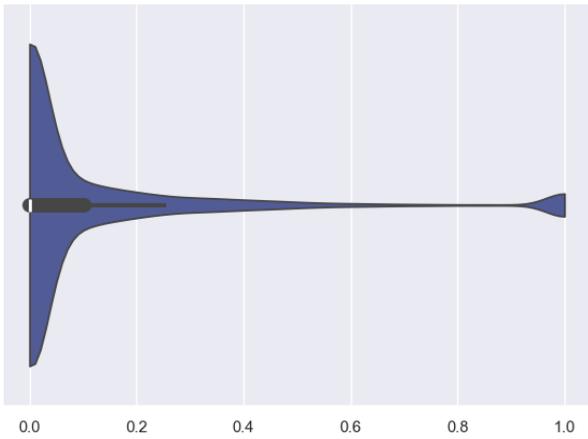


FIGURA A.33. Distribuição da variável "Sistema nervoso e cessação tabágica"

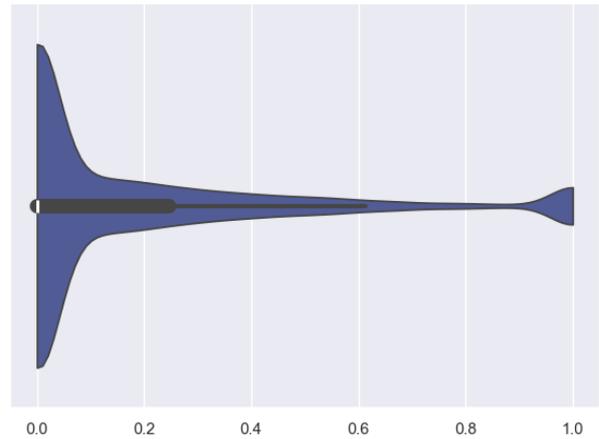


FIGURA A.34. Distribuição da variável "Sistema respiratório"

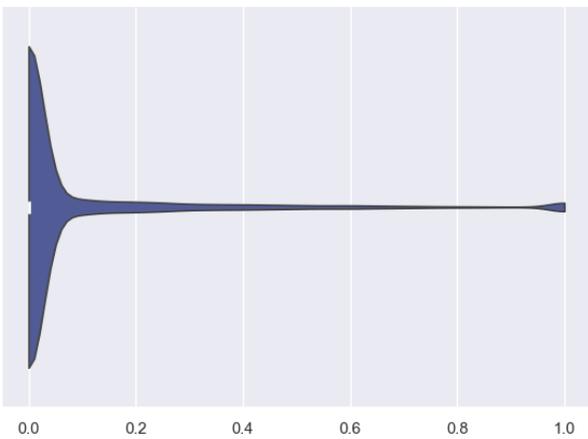


FIGURA A.35. Distribuição da variável "Sistemas musculoesquelético e circulatório"

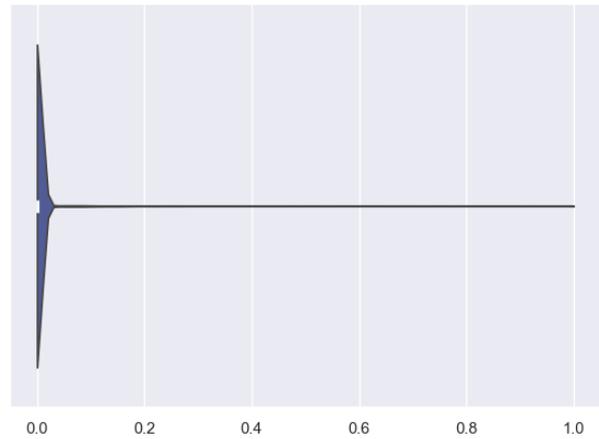


FIGURA A.36. Distribuição da variável "Solares"

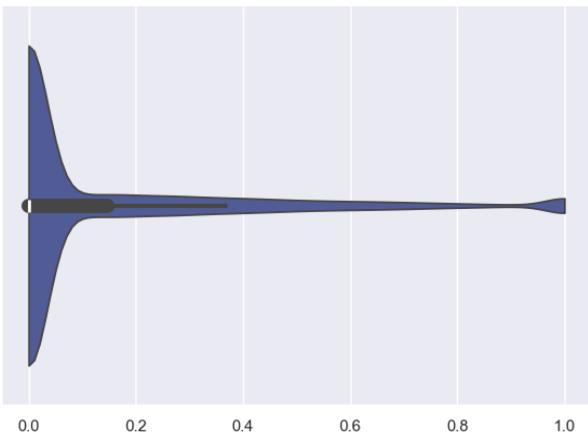


FIGURA A.37. Distribuição da variável "Suplementos alimentares"