

Metodologia de Selecção de Segmentações
Diversificadas:
Um Caso de Aplicação de Técnicas de Data
Mining em Dados de Consumo para Avaliação
de Portfólios de Cartões Bancários

por

Nelson Manuel da Costa Cunha

Dissertação de Mestrado em Gestão Comercial

Orientada por

Prof. Doutor Carlos Manuel Milheiro de Oliveira Pinto Soares

E co-orientada por

Prof. Dr. Alípio Mário Guedes Jorge



***“Learn from yesterday, live for today, hope for tomorrow.
The important thing is not to stop questioning.”
Albert Einstein (1879 - 1955)***

Agradecimentos

Gostaria de agradecer a todos que, de uma forma ou de outra, ajudaram na realização deste trabalho. Em especial ao meu orientador o Professor Doutor Carlos Manuel Milheiro de Oliveira Pinto Soares pela sua paciência, boa disposição, compreensão, conselhos e total disponibilidade que revelou em todas as fases do trabalho e, acima de tudo, pela confiança que depositou em mim.

Ao Prof. Alípio Mário Guedes Jorge, co-orientador, pelo seu conhecimento transmitido sobre o tema, o tempo e dedicação dispensados nas revisões e reuniões.

Queria também agradecer ao Laboratório de Inteligência Artificial e Apoio à Decisão - INESC Porto L.A.

À empresa que disponibilizou os dados para este trabalho.

Aos colaboradores da SAS Portugal pela ajuda prestada e também por me ter disponibilizado uma licença de utilização de software.

À minha família, em especial à minha mãe e ao meu pai, pelo seu apoio e compreensão incondicional desde o primeiro dia.

A todos que me deram apoio para concluir este trabalho.

Nelson Cunha

Areias de Vilar, 2009

Resumo

O tema deste trabalho é a segmentação em *marketing*. Aparentemente a segmentação pode parecer fácil, mas revela-se ser um problema complexo, sobretudo porque frequentemente os objectivos não estão claramente definidos. Seguimos uma abordagem de *Data Mining* que usa o *clustering* para extrair informação útil dos dados e agrupar os mesmos de modo a facilitar a criação e interpretação de perfis de consumo que se assemelhem entre si, a nível psico-demográfico e de comportamento. Tais segmentações podem servir para identificar segmentos com comportamentos e motivações de compra homogéneos para os quais poderá ser dirigida uma oferta específica, como também poderão servir na previsão do perfil de gastos de novos clientes.

Dado que é fácil utilizar ferramentas de *Data Mining* para gerar várias segmentações diferentes, a escolha da segmentação mais adequada surge como um importante problema. Assim propusemos uma metodologia para a selecção (semi-) automática de segmentações (potencialmente) interessantes. A metodologia é empiricamente testado no domínio de aplicação escolhido, o da área bancária. Exemplificámos, como a metodologia pode ser usada para analisar a adequação do portefólio de cartões de crédito e de débito, com base em informações de dois bancos portugueses

Em geral, conclui-se que diferentes segmentações obtidas, usando diferentes condições experimentais (*e.g.*, variando a base de segmentação, o método de *clustering* e/ou o número de clusters) dão perspectivas diferentes sobre o conjunto de clientes analisados. O caso ilustrado também demonstra como a metodologia pode ser usada para identificar cartões de crédito ou de débito que não são adequados para os segmentos de clientes, bem como identificar segmentos para os quais não há oferta adequada de cartões e, nesses casos, apoiar o desenvolvimento de novos produtos adequados para esses segmentos.

Palavras-chave: Data Mining, Marketing bancário, Segmentação, Clustering

Abstract

The scope of this work is segmentation in marketing. Although it may seem easy, it turns out to be a complex problem, because the goals are frequently not clearly specified. We follow a data mining approach, using clustering to extract useful information in the data and group similar customers in order to facilitate the creation and interpretation of consumption profiles, in terms of psycho-demographic and consumption variables. Such segmentations can be used to identify segments with homogeneous behaviors and motivations of purchase, based on which specific offerings could be directed and also to forecast the profile of expenses of new customers.

Given that it is easy to use data mining tools to generate many different segmentations, the choice of the most appropriate segmentation(s) arises as an important problem. We propose a methodology for the (semi-) automatic selection of segmentations, which are (potentially) interesting. The methodology is empirically tested on the chosen domain of application, banking. Additionally, we illustrate how the methodology can be used to analyze the adequacy of the portfolio of credit and debit cards, based on information from two Portuguese banks.

In general, we conclude that different segmentations obtained using different experimental settings (*e.g.*, varying the segmentation base, the method of clustering and the number of clusters), give different perspectives on the set of customers analyzed. Our illustrative study also shows how the methodology can be used to identify credit or debit cards that are not suitable for the customer segments as well as identify segments for which there is no suitable offering of cards and, in those cases, support the development of new, suitable products.

Keywords: Data Mining, Bank Marketing, Segmentation, Clustering

Résumé

Le thème de ce travail est la segmentation en *marketing*. Apparemment, la segmentation peut paraître facile, mais s'avère être un problème complexe, car fréquemment les objectifs ne sont pas clairement définis. Nous avons suivi une approche de *Data Mining* qui utilise le *clustering* pour extraire des informations utiles dans les données et les agrouper afin de faciliter la création et l'interprétation de profil de consommation au niveau psycho-démographique et comportementale. Telles segmentations peuvent servir à identifier les segments avec des comportements et des motivations d'achat homogènes auxquelles une offre spécifique pourra être dirigée, et aussi de prévoir le profil de dépenses de nouveaux clients.

Étant donné qu'il est facile d'utiliser des outils de *Data Mining* pour gérer de nombreuses segmentations différentes, le choix de la segmentation la plus appropriée pose un grand problème. Ainsi nous proposons une méthodologie pour la sélection (semi-) automatique de segmentations (potentiellement) intéressantes. La méthodologie est empiriquement testée dans le domaine d'application choisie qui est celui du domaine bancaire. Nous expliquons également comment cette méthodologie peut être utilisée pour analyser l'adéquation du portefeuille de cartes de crédit et de débit, fondée sur des informations provenant de deux banques portugaises.

En général, on peut conclure que différentes segmentations obtenues en utilisant différentes conditions expérimentales (par ex. en variant la base de segmentation, la méthode de *clustering* et/ou le nombre de clusters) donnent des perspectives différentes sur l'ensemble des clients analysés. Le cas illustré démontre également comment la méthodologie peut être utilisée pour identifier les cartes de crédit ou de débit qui ne conviennent pas pour les segments de clients, ainsi qu'à identifier les segments pour lesquels il n'y a pas offre appropriée et, dans ces cas, soutenir le développement de nouveaux produits adaptés pour ces segments.

Mots-clés : Data Mining, Marketing bancaire, Data Mining, Segmentation, Clustering

Índice

Agradecimentos.....	v
Resumo.....	vii
Abstract	ix
Résumé.....	xi
Lista de Tabelas	xvii
Lista de Figuras.....	xix
Listagem de Siglas.....	xxi
1. Introdução	1
2. Segmentação em Marketing.....	7
2.1. Segmentação na Perspectiva do Negócio	9
2.2. Segmentar: Benefícios & Limitações	11
2.3. Processo de Segmentação	13
2.3.1. Primeira Fase: A Segmentação Propriamente Dita	14
2.3.2. Segunda Fase: o Targeting.....	20
2.3.3. Terceira Fase: O Posicionamento	23
2.3.4. Quarta fase: O Desenvolvimento de um Programa de <i>Marketing (marketing mix)</i>	25
2.4. Áreas de Aplicação da Segmentação.....	27
2.4.1. Segmentação de Mercados na Banca.....	29
3. Data Mining para Segmentação	33
3.1. Conceito de Data Mining.....	34
3.2. Aplicações de Técnicas de Data Mining.....	38
3.3. Metodologias para o Data Mining.....	42
3.3.1. CRISP-DM	42
3.3.2. SEMMA.....	44
3.3.3. CRISP-DM VS SEMMA.....	46
4. Clustering	49

4.1. Definição do Problema de Clustering	50
4.2. Objectivos do Clustering	51
4.3. Aplicações do Clustering	52
4.4. Clustering & Segmentação.....	53
4.5. Métodos de Clustering.....	55
4.5.1. Distância entre Exemplos	57
4.5.2. Método Hierárquico	58
4.5.3. Método não Hierárquico	61
4.5.4. Métodos Fuzzy (difuso)	62
4.6. Avaliação de Métodos de Clustering.....	63
5. Selecção dos Clusterings.....	65
5.1. Definição do Problema	65
5.3. Metodologia de Selecção de Segmentações (Semi-) Automática Simples	66
5.4. Selecção de Segmentações usando Índices de Similaridade.....	69
5.5. Sumário.....	76
6. Resultados: Perspectiva Técnica	79
6.1. Preparação dos Dados	79
6.2. Descrição das Experiências: Métodos e Parâmetros	82
6.2.1. Determinação dos Clusters	82
6.3. Selecção de Segmentações	84
6.4. Descrição e Análise dos Resultados (das Segmentações).....	86
6.4.1. Segmentação W3F	88
6.4.2. Segmentação K3P	89
6.4.3. Segmentação K5P	90
6.4.5. Segmentação H5P	92
6.4.5. Segmentação W5F	93
6.4.6. Segmentação K10P	94
6.4.7. Segmentação W10P	96
6.5. Sumário.....	98
7. Resultados: Perspectiva de Negócio	99
7.1. Cartões Bancários	99

7.2. Cartões Bancários Seleccionados	104
7.3. Alinhamento de Cartões com os Segmentos Obtidos	110
7.3.1. Segmentação K10P	110
7.3.2. Segmentação K5P	115
7.3.3. Segmentação W10P	117
7.4. Desenvolvimento de Novos Cartões	121
7.5. Sumário.....	123
8. Conclusão e Trabalho Futuro.....	125
8.1. Sumário.....	125
8.2. Contributo da Tese	127
8.3. Limitações do Estudo	128
8.4. Trabalho Futuro	129
Referências Bibliográficas.....	131
ANEXOS.....	155
ANEXO A	157
ANEXO B	163
ANEXO C	167
ANEXO D	169
ANEXO E	175

Lista de Tabelas

2.1. - Classificação das bases de segmentação	15
3.1. - Tarefas de Data Mining	36
4.1. - Descrição de termos matemáticos usados no Clustering.....	50
5.1. - Exemplo de Matriz de confusão	67
5.2. - Exemplificação da comparação de clusterings por pares com o mesmo NC	68
5.3. - Exemplo de segmentações semelhantes	68
5.4. - Exemplo de segmentações diferentes.....	68
5.5. - Exemplo de matriz de confusão usado para a comparação dos clusterings	70
5.6. - Matriz de concordância entre os clusterings	71
5.7. - Exemplo de resultados obtidos a partir das expressões matemáticas desenvolvidas por Jain e Dubes (1988) e Albatineh <i>et al.</i> (2006).....	72
5.8. - Exemplo de uma matriz de similaridade	74
5.9. - Resultados com número de clusters de três (NC=3)	74
5.10. - Exemplo de selecção dos clusterings com um diferente NC (5x3).....	75
6.1. - Descrição da escala de idade	80
6.2. - Descrição da variável escala de gastos	80
6.3. - Exemplo de resultados em termos dos volumes de gastos	81
6.4. - Exemplo de resultados em termos de proporção	81
6.5. - Resultados obtidos usando as metodologias desenvolvidas.....	85
6.6. - Valor médio da escala de gastos e idade da amostra.....	86
6.7. - Valores médios e peso nas despesas efectuadas.....	87
6.8. - Segmentação W3F.....	88
6.9. - Segmentação W3F (em proporção).....	88
6.10. - Segmentação K3P.....	89
6.11. - Variáveis com a proporção de despesas significativamente acima da média da segmentação K3P	90

6.12. - Segmentação K5P.....	91
6.13. - Variáveis com a proporção de despesa significativamente acima ou abaixo da média da segmentação K5P.....	91
6.14. - Segmentação H5P.....	92
6.15. - Variáveis com proporção de despesa significativamente acima da média da Segmentação H5P	92
6.16. - Segmentação W5F.....	93
6.17. - Variáveis com valor significativamente acima ou abaixo da média do Cluster 3	94
6.18. - Segmentação K10P.....	95
6.19. - Grau de especialização de cada cluster da segmentação K10P.....	95
6.20. - Segmentação W10P.....	96
6.21. - Grau de especialização de cada cluster da segmentação W10P	97
7.1. - Principais tipos de cartões e o seu funcionamento	102
7.2. - Atribuição de cartões em função de categorias de clientes	103
7.3. - Produtos seleccionados do Santander Totta	104
7.4. - Produtos seleccionados da Caixa Geral de Depósitos.....	106
7.5. - Cartões do S.T. em termos das características usadas nas segmentações	108
7.6. - Cartões do C.G.D. em termos das características usadas nas segmentações	109
7.7. - Identificação de possíveis perfis de consumo com a segmentação K10P.....	110
7.8. - Alinhamento dos cartões do S.T. e da C.G.D. com a Segmentações K10P	115
7.9. - Identificação de possíveis perfis de consumo com a segmentação K5P.....	116
7.10. - Alinhamento dos cartões do S.T. e da C.G.D. com a Segmentação K5P	117
7.11. - Identificação de possíveis perfis de consumo com a segmentação W10P	118
7.12. - Alinhamento dos cartões do S.T. e da C.G.D. com a Segmentação W10P.....	121

Lista de Figuras

2.1. – Níveis de Segmentação de Mercado	9
2.2. – Processo de segmentação de mercado	14
2.3. – Processo de Selecção de Segmentos no mercado B2B.....	23
2.4. – O Marketing Mix	26
3.1. – Fases do processo de KDD/DCBD	35
3.2. – Inquérito sobre as aplicações de Data Mining (por indústria) realizado no mês de Dezembro de 2008) pelo site kdnuggets.com.....	41
3.3. – Estudo sobre a preferência de metodologia usada para o Data Mining.....	42
3.4. – Ciclo de vida da metodologia CRISP-DM	43
3.5. – Software SAS Enterprise Miner 5.2	44
3.6. – Etapas da metodologia SEMMA	46
3.7. – Comparação das interacções entre as fases das metodologias SEMMA e CRISP -DM	48
4.1. – Clustering & Segmentação	53
4.2. – Exemplo de Clustering / Segmentação	54
4.3. – Exemplo de diferentes possibilidade de <i>Clustering</i> a partir do mesmo conjunto dos dados	55
4.4. – Classificação dos algoritmos de clustering	56
4.5. – Exemplificação gráfica da distância euclidiana.....	57
4.6. – Métodos Linkage	59
4.7. – Método do centróide	60
4.8. – Método k-médias	62
5.1. – Processo de selecção semi-automática simples com diferentes pares	69
5.2. – Metodologias de selecção de clustering	77
6.1. – Interface do SAS Enterprise Guide 4.2.....	83
6.2. – Métodos de clustering disponíveis no SAS Enterprise Guide	83

6.3. – Exemplo de determinação de clusters com recurso a ferramenta SAS Enterprise Guide	84
6.4. – Representação gráfica da escala de gastos e idade da amostra	87
6.5. – Distribuição das idades pelos clusters do K10P	96
7.1. – Utilização dos instrumentos de pagamento em PORTUGAL (1992-2002)	100
7.2. – Avaliação dos clientes em função do rendimento e o nível de risco.....	103

Listagem de Siglas

B2B - Business to Business (Marketing)

B2C - Business to Consumer (Marketing)

BI - Business Intelligence

BD – Base de dados

CRISP-DM - **C**Ross-**I**ndustry **S**tandard **P**rocess for **D**ata **M**ining

CRM – Customer Relationship Management

DCBD - Descoberta de Conhecimentos em Bases de Dados

DM - Data Mining

DNA - Deoxyribonucleic Acid

ECD - Extração de Conhecimento de Dados

EM - Enterprise Miner

FV - Força de Vendas

IA - Inteligência Artificial

KDD - Knowledge Discovery in Databases

LTV - Lifetime Value

NC – Número de clusters

O2O - One-to-One (Marketing)

SEMMA - Sample; Explore; Modify; Model; e Assessment

STP - Segmentação, Targeting (Escolha de alvos) e Posicionamento (Marketing)

CAPÍTULO I

*“The way to get started is to quit
talking and begin doing.”*

(Walt Disney)

Introdução

Esta dissertação enquadra-se na problemática da segmentação em marketing e da selecção de segmentações, obtidas com aplicação de técnicas de *Data Mining* (DM).

Actualmente, nos mercados, o sucesso já não depende da capacidade da Força de Vendas (FV) em comunicar o valor do produto ou dos serviços (Justino, 2007), mas da capacidade das empresas de criarem valor aos seus clientes. Para isso, é necessário “um conhecimento profundo do cliente e das suas necessidades” (*idem*).

Para entender melhor as necessidades e desejos dos seus clientes é necessário atendê-los de forma mais eficiente, contudo, para a maioria das empresas torna-se muito complicado, a nível financeiro, satisfazer estas necessidades individualmente (*Marketing One-to-One*).

Perante a dificuldade em satisfazer eficazmente às necessidades de todos os seus clientes, uma das soluções apresentadas é o processo de segmentação de mercado (Cant *et al.*, 2009), que permite agrupar os clientes em grupos relativamente homogéneos. A prática da segmentação tornou-se incontornável para que as empresas consigam concentrar os esforços de *marketing* em determinados alvos (*Target*) considerados como “favoráveis” para serem explorados comercialmente.

Para a concretização da segmentação, é necessário que as empresas detenham um conhecimento mais apurado das características dos seus clientes, assim sendo, é necessário recolher dados sobre as interações entre os clientes e a empresa a fim de determinar um perfil dos seus clientes. No entanto, grande parte deste conhecimento útil está escondido e inexplorado (Shaw *et al.*, 2001), devido ao crescente volume dos dados

que dificulta a análise (manual) pelos peritos de marketing. A fim de solucionar este problema, surgiram ferramentas computacionais, com destaque para o *Data Mining*, que permitem extrair informações a partir de dados históricos. O objectivo é obter mais conhecimento sobre as tendências de compras dos clientes e usá-las para a tomada de decisões e para a realização de previsões. O *Data Mining* é mesmo considerado, por muitos autores (e.g. Berry e Linoff, 2004), como uma ferramenta incontornável num serviço de *marketing* e vendas, e no seio dos processos de decisão da empresa.

A segmentação é uma abordagem analítica clássica que permite estudar qualquer assunto difícil de compreender na sua globalidade, quer seja filosófico, científico ou comercial. A segmentação permite “dividir os dados de uma forma conveniente” (Collica, 2007) e analisá-los sob os seus diferentes aspectos, de dissociar os seus constituintes, de estudá-lo um por um antes de apreender a síntese. O *marketing* apropriou-se deste método analítico para o estudo dos mercados (Kotler e Dubois, 2000). No entanto, devido à complexidade envolvida na sua realização, é, hoje em dia, considerado por vários autores como uma arte. Adicionalmente, a segmentação por si só não é suficiente para explicar tudo, mas é uma fase essencial de qualquer abordagem de *marketing* (Kotler e Keller, 2006).

Na prática, a maioria dos estudos de segmentação comercial envolve a análise de clusters (*clustering*) (Hoeck *et al.*, 1998), porque permite agrupar automaticamente os dados de forma a identificar segmentos (grupos) com características homogéneas (Everitt, 1993). Assim, os grupos criados facilitam a interpretação dos perfis de consumo existentes, permitindo elaborar estratégias mais adaptadas para satisfazer as necessidades e desejos dos clientes, o que torna possível a identificação de oportunidades de lançamento de novos produtos e melhorar os produtos e serviços prestados existentes.

Uma das áreas de negócio com maior investimento nos processos de segmentação (de clientes) é a área bancária que considera a segmentação “como uma importante ferramenta de apoio à gestão empresarial” (Kaynak e Harcar, 2005). Nos últimos anos, em razão de uma variedade de factores, como a liberalização e a globalização, a banca confrontou-se com uma maior concorrência, quer a nível interno, quer externo. Para serem competitivos num ambiente de mercado volátil, os bancos recorrem a ferramentas e técnicas de *marketing* (*idem*) para compreenderem melhor as

necessidades dos clientes. O objectivo final é o de tentarem atrair (potenciais) novos clientes e manterem os actuais através da promessa de um melhor serviço ao cliente e de uma oferta de produtos mais completa relativamente aos da concorrência. O grande desafio dos bancos passa, hoje, por grandes mudanças na aplicação de técnicas de *marketing*, como a segmentação dos clientes e o desenvolvimento de acções comerciais.

O presente trabalho é baseado em dados psico-demográficos e de consumo de clientes de uma empresa. As segmentações criadas com base neste tipo de dados, normalmente têm por objectivos:

- (i) Criar grupos de clientes que se assemelhem entre si, a nível psico-demográfico, como a nível de consumo.
- (ii) Identificar segmentos com comportamentos e motivações de compra homogéneos, aos quais poderá ser dirigida uma oferta específica;
- (iii) Fornecer uma segmentação de clientes em termos de gastos para servir na previsão do perfil de gastos de novos clientes;

Se, por um lado, técnicas como o *Data Mining* abriram novas oportunidades em áreas como o *marketing* (e.g., segmentação automática de mercados), por outro, estas novas aplicações abriram novas perspectivas para a área da investigação académica. Por exemplo, a facilidade de utilização de técnicas de DM para *marketing* levanta alguns novos desafios. Deste modo, é possível, hoje em dia, fazer rápida e facilmente um grande número de segmentações diferentes. Analisar manualmente todas as segmentações geradas é, pois, impraticável.

Assim, este trabalho tem por objectivos:

- (i) Desenvolver um método que permita seleccionar um pequeno número de segmentações, de um grande conjunto gerado automaticamente;
- (ii) Testar esse método nos dados mencionados anteriormente;
- (iii) Ilustrar a utilidade do método, tanto para análise da adequação dos portfólios de cartões de crédito e débito de duas instituições bancárias portuguesas aos segmentos de clientes identificados, como para o apoio

ao desenvolvimento de novos produtos para segmentos de clientes para os quais não existe oferta adequada.

A dissertação encontra-se organizada em duas partes. Na primeira parte, são apresentados os principais fundamentos teóricos associados à temática da segmentação, *Data Mining* e ao *clustering*. Os capítulos de revisão teórica procuram sintetizar e discutir alguns conceitos considerados relevantes para a investigação. Na segunda parte, procede-se à investigação empírica sobre a selecção das segmentações, sendo que o âmbito de aplicação do estudo empírico é a área bancária onde se vai tentar alinhar os produtos bancários (*i.e.* cartões bancários) com as segmentações seleccionadas.

Relativamente à primeira parte, a dissertação estrutura-se em quatro capítulos. No capítulo I, procede-se à introdução sobre a natureza e a complexidade associadas à segmentação.

O capítulo II debruça-se sobre a apresentação da problemática da segmentação em *marketing*, com destaque para o processo de segmentação (**Segmentação, Targeting, Posicionamento**). Neste capítulo, tentar-se-á olhar para as várias linhas de trabalho em segmentação e olhar para o problema como um todo.

No terceiro capítulo, é efectuada uma apresentação dos conceitos fundamentais sobre o *Data Mining* para segmentação e respectivas técnicas que surgem como a principal alternativa para a análise e exploração de uma grande quantidade de dados, de difícil resolução através de abordagens mais convencionais.

A revisão da literatura encerra com o capítulo IV, onde é apresentado um estudo amplo do problema de *clustering*, tendo como principal enfoque a análise de *clusters* no *marketing*, destacando as suas principais características e conceitos.

A parte empírica da dissertação inicia com a apresentação da metodologia desenvolvida para a selecção dos *clusterings* (capítulo V). Neste capítulo, assumimos que é criado um grande número de diferentes segmentações da base de dados usando diferentes algoritmos de *clustering*, discutidos no capítulo IV. Assim sendo, são apresentados métodos (semi-) automáticos que foram desenvolvidos para a selecção de segmentações (potencialmente) interessantes.

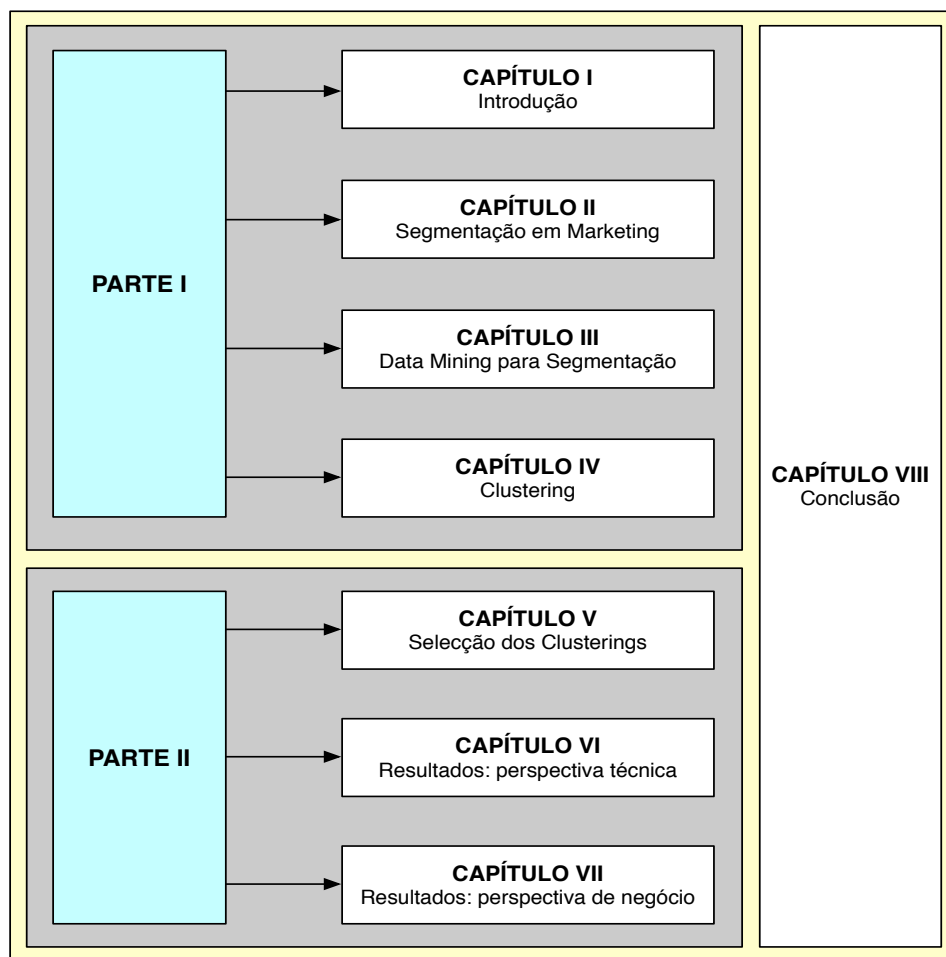
No capítulo VI, descrevem-se os resultados das segmentações geradas de forma detalhada numa perspectiva técnica. São também descritos os dados que foram usados e

as transformações realizadas nos mesmos. Neste capítulo, vamos ainda apresentar e descrever as segmentações seleccionadas usando os métodos propostos no capítulo anterior.

No capítulo VII, faz-se uma síntese dos resultados na perspectiva de negócio. Mais concretamente, vamos analisar e descrever a oferta existente no mercado nacional de produtos (*i.e.* cartões bancários) de duas entidades bancárias e tentar alinhar os respectivos produtos com as segmentações seleccionadas que foram apresentadas no capítulo VI. É ainda ilustrado como podem ser criados novos produtos, especificamente para segmentos que não são cobertos (ou apenas parcialmente) pela oferta existente.

O capítulo VIII encerra a dissertação com a apresentação das conclusões e o seu contributo para uma melhor compreensão do fenómeno em estudo. São ainda debatidas algumas das limitações inerentes ao estudo e, para concluir, propõem-se algumas pistas para investigações futuras.

Esquemáticamente, a estrutura da dissertação é a seguinte:



CAPÍTULO II

“The fact is, everyone is in sales. Whatever area you work in, you do have clients and you do need to sell.”

Jay Abraham in *Getting Everything You Can Out of All You've Got* (2001:6)

Segmentação em Marketing

A heterogeneidade das necessidades dos clientes, nos mercados, torna muito difícil às empresas poder desenvolver um pequeno conjunto de produtos/serviços que os satisfaçam a todos. Tentar fazê-lo conduz, geralmente, a uma diminuição da rentabilidade. Por isso, é necessário dividir os clientes em subgrupos homogêneos. Esta abordagem foi denominada *segmentação de mercado* pelo autor Wendell Smith (1956), que impulsionou desenvolvimentos ao nível teórico e prático (Wedel e Kamakura, 1999)¹.

A segmentação de mercado pode ser descrita mais concretamente como o processo de divisão de um grande mercado em mercados alvo, mais pequenos, ou *cluster* de clientes (Smith, 1956; Kotler e Gordon, 1983; Croft, 1994; Myers, 1996).

O objectivo de uma segmentação é dividir o mercado de massa em submercados de clientes que têm necessidades comuns. Identificar esses segmentos torna possível fazer duas coisas:

- a) criar bens e serviços que são mais adequados às necessidades específicas dos clientes;
- b) focar os recursos de *marketing* mais eficientemente (*Harvard Business Essentials*, 2006).

Basicamente, a segmentação tenta concentrar os esforços de *marketing* em

¹ O interesse académico em termos de investigação sobre a segmentação é ilustrado pelas 1600 referências identificadas por Wedel e Kamakura (1999) numa pesquisa realizada sobre a literatura de *marketing*.

determinados alvos que a empresa entende como favoráveis para serem explorados comercialmente, em decorrência da sua capacidade de satisfazer a procura, e isso de maneira mais adequada.

Ao longo dos últimos 50 anos, houve um reconhecimento generalizado por parte de académicos e profissionais quanto à importância e aos benefícios da segmentação, quer para o consumidor, quer para as empresas, não só em contexto de *Business-To-Consumer* (B2C) como de *Business-To-Business*² (B2B).

A segmentação é frequentemente diferenciada, entre os académicos que a discutem, como uma actividade estratégica ligada à perseguição das vantagens competitivas (Smith, 1956; Freytag e Clarke, 2001; Kotler *et al.*, 2005;). Um dos problemas fundamentais da literatura sobre a segmentação é que “reflecte um afastamento entre os académicos e as necessidades dos práticos” (Dibb, 1998:397). Dibb (1998) salienta ainda que “o uso de técnicas e linguagem complexa é uma crítica particularmente comum no que diz respeito a literatura sobre segmentação”.

Goller *et al.* (2002) classificou a investigação académica em segmentação, em quatro áreas principais:

- a) *O desenvolvimento das bases de segmentação e modelos* (e.g. Haley, 1968; Wind e Cardozo, 1974; Moriarty e Reibstein, 1986);
- b) *Pesquisas de metodologias* (e.g. Kalwani e Silk, 1982; Flodhammar, 1988);
- c) *O desenvolvimento e aplicação de ferramentas de análise estatística* (e.g. Rao e Winter, 1977; Green e Krieger, 1991; Balakrishnan *et al.*, 1996);
- d) *Implementação da segmentação* (e.g. Mahajan e Jain, 1978; Piercy e Morgan, 1995).

Os mesmos autores (Goller *et al.*, 2002) afirmam que as áreas de investigação acima referidas se desenvolveram essencialmente em isolamento, resultando numa compreensão fragmentada do processo como um todo. O que, no início, parece ser um processo relativamente “clássico” de Segmentação, *Targeting* e Posicionamento (STP) (Kotler *et al.*, 2005), na realidade, torna-se um campo de estudo altamente complexo.

² Também designado como mercado organizacional ou industrial.

De seguida, discutir-se-á a Segmentação, numa perspectiva do negócio (Secção 2.1). Depois, abordar-se-á o processo de segmentação (Secção 2.2). Finalmente, descrever-se-ão algumas áreas de aplicação da segmentação (Secção 2.3), em particular, no âmbito da Banca (Secção 2.3.1).

2.1. Segmentação na Perspectiva do Negócio

A segmentação tornou-se um conceito de ordem estratégica mais do que operacional (Tassel, 2004) e deveria ser mesmo, de acordo com Hooley *et al.* (1998), o primeiro passo na definição de qualquer estratégia de *marketing*. Por segmentação, referimo-nos ao processo no qual os gestores de *marketing* tentam perceber o mercado, recolhendo e analisando diversas variáveis, recorrendo a várias técnicas sofisticadas (*e.g.* análise de clusters) (Hoek *et al.*, 1996).

De acordo com Kotler e Keller (2006) e Oliveira-Brochado e Martins (2008), a segmentação de mercado pode ser realizada em vários níveis (ver figura 2.1.), desde a ausência de segmentação (*marketing* de massas) até à segmentação completa (micro segmentação), ou com recurso a abordagens intermédias (*marketing* por segmentos ou *marketing* por nichos³).

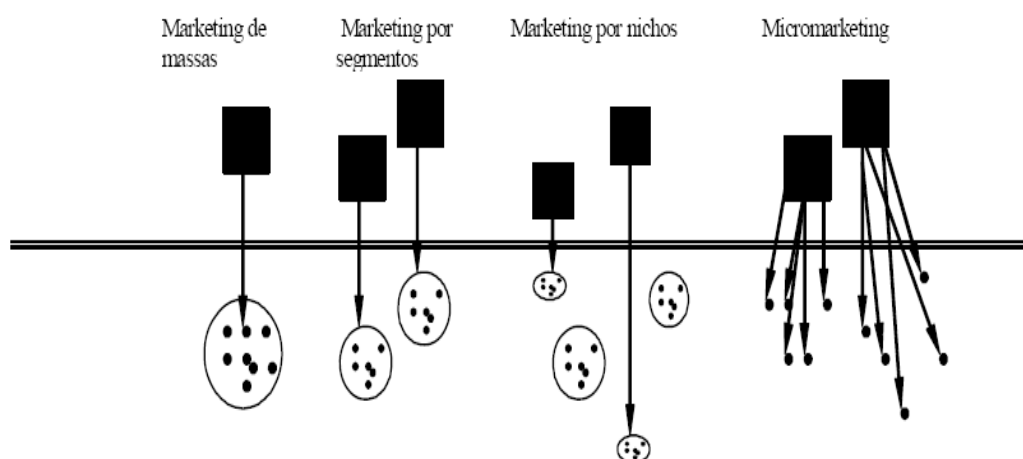


Figura 2.1. – Níveis de Segmentação de Mercado (Oliveira-Brochado e Martins, 2008)

³ Um Nicho de Mercado corresponde a um segmento de mercado constituído por um reduzido número de consumidores com características e necessidades homogêneas e facilmente identificáveis. Os gestores de *marketing* costumam identificar nichos (um nicho é mais pequeno que um segmento), dividindo um segmento em subsegmentos.

No *marketing de massas*, a empresa produz, distribui e promove em massa um produto para todos os consumidores. De acordo com Kotler e Keller (2006), o *marketing de massas* conduz a custos mais reduzidos, preços mais baixos ou margens mais altas e abre as portas a mercados mais vastos. No entanto, muitos críticos (ver Kotler e Keller, 2006) apontam para o aumento da divisão do Mercado, o que torna o *marketing de massas* mais difícil.

Em relação ao *marketing por segmentos*, a empresa identifica diversos segmentos no mercado, escolhe um ou mais, e desenvolve produtos e compostos de *marketing* ajustados a cada segmento. Os consumidores procuram variedade e mudança, e apresentam gostos diferentes que mudam com o tempo. Segundo Kotler e Keller (2006), o *marketing por segmentos* oferece vantagens sobre o *marketing de massas*. A empresa pode presumivelmente melhorar “o design, o preço, divulgar e entregar o produto ou serviço para satisfazer o mercado-alvo” (*idem*).

No que diz respeito ao *marketing por nichos*, devido à sua pequena dimensão, os nichos de mercado são geralmente desprezados pelas grandes empresas, constituindo, por isso, excelentes oportunidades para as pequenas empresas que aqui podem conseguir uma posição de liderança através de uma oferta muito específica e adaptada às características e necessidades dos consumidores que constituem o nicho (Nunes, 2007). De acordo com Kotler *et al.* (2005), uma das vantagens dos nichos sobre o *marketing por segmentos* é que os nichos atraem um ou dois concorrentes, mas também aumentam a exigência do consumidor, e a diferenciação dos produtos e serviços deve ser muito maior, pois o mercado inserido num nicho é muito menor, e o público-alvo conhece mais profundamente o produto, tornando fácil a substituição de uma empresa por outra. Para que isso não ocorra, deve-se ter um trabalho de fidelização muito bem elaborado e eficiente. Em relação ao *micromarketing* (ou *Marketing One-to-One*⁴), a empresa ajusta os produtos e os seus programas de *marketing* às necessidades e desejos de um pequeno e bem definido segmento, que, em caso extremo, é um único consumidor (Kotler *et al.*, 2005). Ou seja, o *Marketing One-to-One* (O2O) pretende satisfazer as necessidades do consumidor, através de um serviço personalizado que

⁴ O *Marketing One-to-One* (ou *marketing individualizado*) significa relacionar-se individualmente com os clientes, um de cada vez.

oferece os produtos adequados ao seu perfil, assim como pretende ajudar a estabelecer uma relação de confiança e lealdade com a empresa. Apesar de tudo, para uma tal abordagem, o facto de tratar cada indivíduo / cliente (perfil, acções, intenções, ...) – torna-se material e financeiramente não rentável para muitas empresas, devido, por exemplo, à importância do volume de clientes e do investimento em tecnologias de informação que a estratégia do O2O exige.

2.2. Segmentar: Benefícios & Limitações

O problema básico relacionado com a segmentação pode parecer relativamente “simples” (Wedel e Kamakura, 1999): a classificação de consumidores em grupos que exibam homogeneidade interna e heterogeneidade entre si (Everitt, 1993). No entanto, é um problema complexo, cuja eficácia depende das variáveis de agrupamento, da metodologia de análise dos dados que foi escolhido e da correcta interpretação dos resultados (Wedel e Kamakura, 1999).

Segundo vários autores (*e.g.* Everitt, 1993; Wedel e Kamakura, 1999), a segmentação é talvez a decisão mais difícil e complexa que uma empresa tem de tomar ao decidir a sua estratégia de mercado, porque uma “correcta segmentação de mercado” pode definir o êxito ou o fracasso comercial da empresa, ou seja, a abordagem da segmentação tem de ser feita com muita reflexão e com cuidado (*Harvard Business Essentials*, 2006). O facto de haver muitas maneiras de segmentar mercados torna essa tarefa ainda mais complicada. É por isso que vários autores se referem à segmentação como uma arte, dado que “definir os segmentos é como criar um mosaico onde muitas peças diferentes são combinados para criar uma lógica atraente de segmento de mercado” (Beine, 2003). A segmentação, para além de ser o caminho mais viável para a subsistência de qualquer organização, oferece inúmeras vantagens. Estas notam-se por exemplo ao nível do consumidor, em relação à concorrência, uma vez que a estratégia a adoptar pressupõe que conheçamos bem as necessidades e características dos consumidores, o que nos torna capazes de obter uma vantagem competitiva em relação aos demais concorrentes.

Os benefícios da segmentação em *marketing* são, por vezes, subestimados. Por exemplo, o processo pode incrementar a competitividade que se traduz em vantagens tais como:

- a) fornecer uma base racional para a selecção dos mercados-alvos⁵;
- b) favorecer maior proximidade ao consumidor final (Richers e Lima, 1991);
- c) desenhar uma oferta de forma adequada às necessidades, desejos, atitudes e poder de compra dos clientes (Kotler *et al.*, 2005; Doyle, 2000);
- d) focar o esforço de distribuição e de comunicação (na esperança de que uma maior satisfação dos clientes conduza a fidelização e a recompra);
- e) ajudar os profissionais de *marketing* a identificar mais facilmente oportunidades de crescimento e ameaças (Bannon, 2003);
- f) identificar oportunidades para o desenvolvimento de novos produtos;
- g) identificar oportunidades para a diferenciação de estratégia (por exemplo, os mercados servidos e o posicionamento do produto);
- h) a personalização da relação com os clientes (*i.e.* CRM) (Seggern e Hadaway, 2004);
- i) dar uma resposta rápida às alterações nas necessidades do mercado: ao conhecermos melhor os subgrupos, especializamo-nos, podendo assim, responder mais rapidamente às necessidades dos clientes-alvo;
- j) uma melhor alocação efectiva de recursos: quanto melhor se conhecer os seus clientes, melhor serão aplicados os recursos existentes (consegue-se responder às necessidades de forma mais eficaz e com maior sucesso) (Kotler *et al.*, 2005).

Apesar das suas vantagens, o processo de segmentação em *marketing* não é isento de críticas. Há várias questões práticas e evidentes com o processo tais como:

- a) a adequação do mercado para aplicar o processo de segmentação: “Uma ênfase excessiva nas técnicas de segmentação pode negligenciar o próprio mercado e a posição competitiva de uma organização nesse mercado” (Bliss, 1980 *apud* Bannon, 2003);

⁵ Entende-se por *mercado-alvo* o segmento específico de mercado que uma organização escolhe para atender.

- b) os custos da análise dos benefícios;
- c) a estabilidade de segmentos produzidos;
- d) a validade e fiabilidade dos critérios para a segmentação: “Qualquer tentativa de utilizar uma única base de segmentação para todas as decisões de *marketing* pode resultar em resultados inadequados e desperdício de recursos” (Wind, 1978);
- e) a avaliação e a interpretação adequada dos dados (Wedel e Kamakura, 1999).

Após a revisão da literatura, constatamos que a segmentação foi extensamente utilizada no *marketing* nos últimos trinta anos, tornando-se um dos principais conceitos de *marketing* (teórico e prático), baseado no pressuposto de que os clientes demonstram preferências heterogêneas e comportamentos de compra (Assael e Roscoe, 1976; Green, 1977; Wind, 1978). Essa heterogeneidade pode, geralmente, ser explicada pelas diferenças no produto e pelas características do utilizador (Blattberg e Sen, 1976; Kalwani e Morrison, 1977; Mahajan e Jain, 1978).

2.3. Processo de Segmentação

De acordo com Kotler (2003), as principais fases do processo de segmentação de mercado são a segmentação propriamente dita, o *Targeting* – escolha de alvos e o posicionamento (S.T.P.).

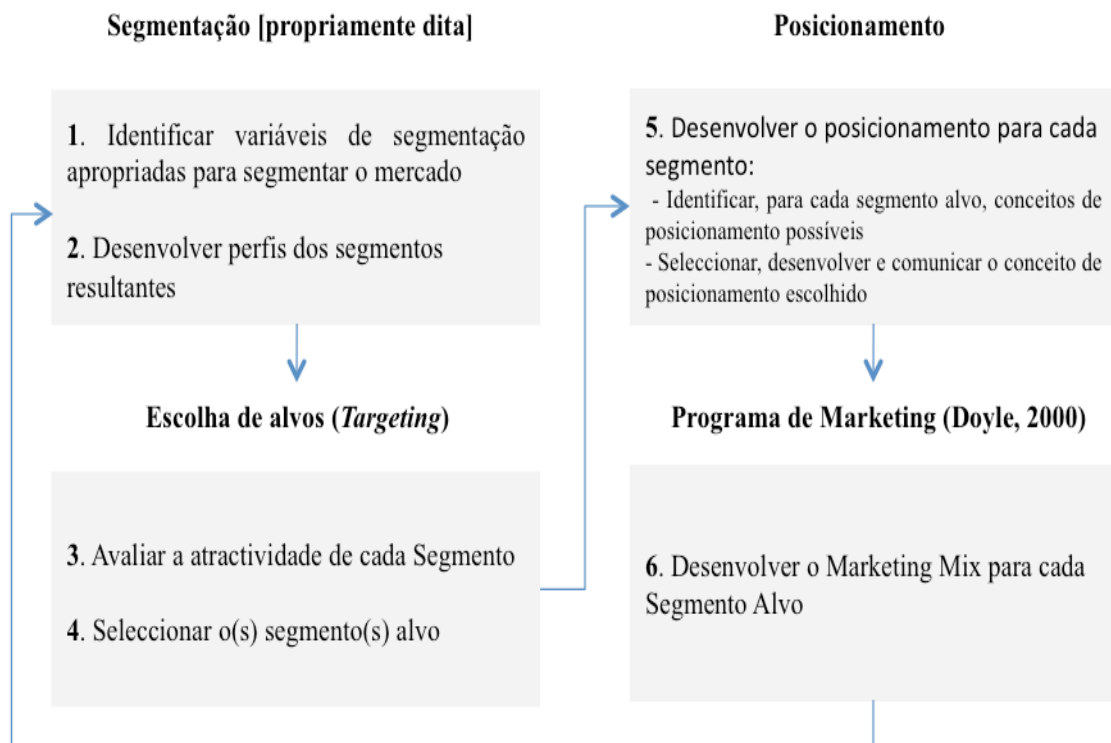
Neste trabalho, seguimos a perspectiva de Doyle (2000), que considera que a estas fases se segue o desenvolvimento e implementação de um Programa de *Marketing* (*marketing mix*), que corresponde à oferta propriamente dita.

Como se pode observar na figura 2.2., estas fases podem ainda ser subdivididas em várias etapas (Pride e Ferrell, 1997; Best, 2000; Doyle, 2000; Kotler *et al.*, 2005):

- 1) identificar bases (critérios) de segmentação apropriadas para segmentar o mercado;
- 2) desenvolver o perfil dos segmentos encontrados;

- 3) avaliar a atractividade de cada segmento;
- 4) seleccionar os segmentos alvos;
- 5) desenvolver o posicionamento para cada segmento;
- 6) e desenvolver o *marketing mix* para cada segmento.

Figura 2.2. – Processo de segmentação de mercado



Feedback obtido do mercado permite o ajustamento da estratégia comercial – alteração do posicionamento, entrada ou saída de segmentos

Adaptado de Brito, 1998; Doyle, 2000; Bannon, 2003; Kotler *et al.*, 2005.

Salientamos o facto de que o processo de segmentação pode ser efectuado de modo manual (baseado no conhecimento do próprio analista) ou (semi-) automático, recorrendo, por exemplo, a técnicas de *Data Mining* (DM) ou de análise multi-variada de dados, conforme será discutido no capítulo III.

2.3.1. Primeira Fase: A Segmentação Propriamente Dita

Esta fase consiste naquilo que podemos definir como a segmentação propriamente dita, ou seja, o agrupamento dos clientes em grupos homogêneos.

A segmentação pode ser dividida nas seguintes etapas: 1) identificar bases (critérios) de segmentação apropriadas para segmentar o mercado; e 2) desenvolver o perfil dos segmentos encontrados.

A primeira etapa do processo de segmentação diz respeito a identificação de variáveis (critérios) para a segmentação de mercado. De acordo com Oliveira-Brochado e Martins (2008), para a identificação de grupos homogêneos, o primeiro passo recai na selecção das bases de segmentação. A base de segmentação diz respeito à natureza dos factores de homogeneidade que caracterizam um grupo de consumidores (Vaz, 2003). Numa aplicação particular, a sua escolha é função de dois elementos: o propósito do estudo (por exemplo, desenvolvimento de um novo produto, selecção dos *media*, fixação do preço), e o mercado em análise (por exemplo, industrial, do consumidor, internacional) (Oliveira-Brochado e Martins, 2008).

Frank *et al.* (1972) sugeriram um sistema de classificação, sendo depois revisto por Wedel e Kamakura (1999), que consiste na classificação das bases de segmentação em função de duas características: a natureza da variável (geral ou específica do produto) e a natureza do processo de medida (observável e não observável), descritas na tabela 2.1..

Tabela 2.1. – Classificação das bases de segmentação

		Natureza da variável	
		Gerais	Específicas do produto
Natureza do processo de medida	Observáveis	Variáveis geográficas, demográficas e socioeconómicas	Ocasões de uso, estatuto de uso, fidelidade, processo de difusão e processamento de informação
	Não observáveis	Variáveis psicográficas, valores, personalidades e estilo de vida	Benefícios, percepções, elasticidades, preferências, intenções, psicográficas (específicas do produto)

Adaptado de Wedel e Kamakura (1999); Oliveira-Brochado e Martins (2008)

No caso concreto do mercado do consumidor, Kotler (2003) sugere que o mercado do consumidor pode ser segmentado com as características do cliente e propõe

quatro grandes tipos de critérios (base de segmentação): *i) geográficos* (região: por continente, país, estado e mesmo vizinhança, tamanho da área metropolitana, densidade populacional e clima); *ii) demográficos* (idade, género, dimensão do agregado familiar, rendimentos, educação, religião, etc.); *iii) psicográficos* (actividades, interesses, opiniões, atitudes, valores, estilo de vida, ...); e *iv) comportamentais / de comportamento* face ao produto (fidelidade da marca, taxa de utilização, benefícios procurados, prontidão para comprar, etc. ...).

Quanto a mercados de bens industriais, dada a especificidade, quer dos produtos, quer dos actores envolvidos, é usual entrar em linha de conta com aspectos relativos ao processo de compra e à própria estratégia das empresas (Brito, 1998). Assim, por exemplo, Zikmund e d'Amico (1993) sugerem uma segmentação com base em critérios *organizacionais* (dimensão da empresa, tecnologia usada,...), *estratégicos* (estratégia competitiva adoptada,...) e *relativos ao processo de compra* (dimensão das encomendas, nível de centralização/descentralização da decisão de compra,...) para além dos de natureza geográfica.

É preciso, no entanto, referir que um estudo de segmentação de mercado depende, não só da base de segmentação e das suas propriedades, mas também do método de segmentação (Oliveira-Brochado e Martins, 2008). Em termos da sua operacionalização, os métodos de segmentação podem ser classificados de acordo com o momento de determinação dos segmentos e com o método estatístico utilizado (Wedel e Kamakura, 1999):

A) Momento de determinação dos segmentos:

A1) A priori, o número de segmentos é definido pelo investigador antes de iniciar a pesquisa, ou seja, os critérios de segmentação estabelecidos antes dos descritores dos segmentos (por exemplo, taxas de uso, fidelidade). Após a definição da base de segmentação, procede-se à descrição dos segmentos (Carmichael, 1996). A escolha da(s) variável(eis) resulta do que o investigador considerar o que é importante para compreender o comportamento do consumidor.

A2) *Post hoc*, o número de segmentos é determinado com base na análise dos resultados da pesquisa, ou seja, procuram-se segmentos significativos sem ter a certeza de quais serão em avanço (utiliza a análise conjunta para identificar grupos de consumidores que possam ser idênticos no que diz respeito a algumas das variáveis). Começa-se pela definição de características importantes, as quais são utilizadas para a constituição de grupos (*clusters*) homogêneos.

B) Método estatístico utilizado:

B1) *Descritivo* – consiste na análise de associações entre um conjunto de variáveis (bases de segmentação), sem que as mesmas sejam classificadas em variáveis dependentes ou independentes⁶.

B2) *Preditivo* – a análise de associações é realizada entre dois conjuntos de variáveis, em que através das independentes (ou explicativas) se procura explicar as dependentes.

Relativamente às abordagens *a priori* e *post hoc* (Green, 1977), a opção por uma base *a priori* pode representar uma antecipação de decisões em relação à resposta do consumidor e ao *marketing mix*. Quer isto dizer que a investigação tem apenas por objectivo a caracterização dos segmentos, pois a sua identificação está assumida desde o início. Por sua vez, as bases *post hoc* podem ter uma maior afinidade com os padrões identificados pela pesquisa, mas não geram necessariamente segmentos com consistência interna (Hoek *et al.*, 1996).

O segundo passo consiste em desenvolver os perfis dos segmentos encontrados, sendo que a definição dos perfis dos clientes (características, motivações, comportamentos, factores condicionantes, influências e locais de compra) é uma etapa importante para melhorar a segmentação (Seggern e Hadaway, 2004). Por perfil, referimos a descrição do consumidor ou de qualquer público-alvo segundo elementos predefinidos. Para que seja possível tentarmos definir o perfil, é necessário

⁶ As variáveis dependentes (endógenas) são aquelas cujos resultados dependem do comportamento das variáveis independentes ou explicativas (exógenas), que são manipuladas e controladas pelo investigador.

conseguirmos identificá-lo de modo a compor um quadro do comportamento de compra. Para conseguir definir um perfil completo do consumidor é necessário considerar simultaneamente várias bases (Oliveira-Brochado e Martins, 2008).

Os métodos utilizados em *marketing* para determinar os perfis dos segmentos podem ser:

- a) *geográficos* (Kelly e Nankervis, 2001; Kotler *et. al.*, 2005) – O mercado é dividido em diferentes unidades geográficas, tais como nação, estados, regiões, municípios, cidade ou proximidades; com esta abordagem, a empresa reconhece que os potenciais de mercado e os custos variam com a localização dos mercados;
- b) *demográficos* (Blattberg *et al.*, 1976; Kelly e Nankervis, 2001) – Os críticos argumentam que um dos maiores problemas da segmentação demográfica é a falta de “*riqueza*” (cor, textura e dimensionalidade) na descrição dos consumidores para a segmentação do mercado e desenvolvimento estratégico, sendo que frequentemente é necessário completá-la para preencher o “*bare statistical picture*”;
- c) *psicográficos* (Alpert, 1972; Frank *et al.*, 1972; Wells, 1975; Zotti, 1985) – De acordo com Demby (1974), a segmentação psicográfica pode ser vista como o método de definir o estilo de vida em termos mensuráveis, sendo definido como “A utilização de factores psicológicos, sociológicos e antropológicos (...), para determinar como o mercado é segmentado pela propensão dos grupos dentro do mercado e da razão para a tomada de uma certa decisão sobre um produto, pessoa, ideologia, ou de outra maneira manter a atitude ou usar um meio.” (Demby, 1974);
- d) *benefícios* (“benefits”) (Haley, 1968; Myers, 1976; Calatone e Sawyer, 1978; Moriarty e Reibstein, 1986; Soutar e McNeil, 1991) – Na segmentação de benefícios, os potenciais consumidores são agrupados de acordo com “a sua utilidade desejada ou prevista de consumir um produto” (Haley, 1968). E demonstrou ser útil no mercado de consumo (Haley, 1968; Myers, 1976), no mercado de bens industriais (Moriarty e Reibstein, 1986) e também no mercado

de bens e serviços (Calantone e Sawyer, 1978; Soutar e McNeil, 1991); o interesse deste tipo de segmentação resulta do facto de fornecer uma perspectiva diferente sobre o mercado. Tal como a segmentação psicográfica, a segmentação por benefícios é apresentada como uma abordagem que permite a identificação de segmentos a partir de factores causais, por oposição aos factores descritivos (Haley, 1968). Embora tenha por base um conceito simples, a sua operacionalidade é complexa, exigindo a aplicação de métodos multivariados a dados obtidos a partir de inquéritos (Haley, 1968);

- e) *utilização* (Twedt, 1964; Young *et al.*, 1978; Bowen, 1998) – A segmentação por utilização é reconhecida como uma das importantes bases para a segmentação num número de mercados de produtos / serviços (Twedt, 1964; Pride e Ferrell, 1989; Kotler, 2003). Esta abordagem trata de dividir o mercado de acordo com a taxa de utilização de um produto e usualmente divide os consumidores em não-utilizador, utilizador leve, médio e pesado. Este método pode, por exemplo, ajudar as empresas a atingirem o objectivo de servir todos os clientes com um processo correspondente às necessidades e capacidade de pagamento;
- f) *estilo de vida* (Lazer, 1963; Yankelovich, 1964; Plummer, 1974) – A premissa base da investigação sobre o *estilo de vida* considera que um melhor conhecimento do consumidor contribui para que o processo de troca e a comunicação sejam mais efectivos. O padrão de estilo de vida combina as virtudes demográficas com a riqueza e a dimensão das características psicológicas (Plummer, 1974);
- g) *fidelidade* (Grover e Srinivasan, 1989; Yelkur e Da Costa, 2001) – A segmentação por fidelidade envolve a identificação da fidelidade dos clientes com uma marca ou produto. Com esta abordagem, os clientes tendem a ser muito leais, moderadamente leais ou desleais. A fidelidade é uma importante variável para a segmentação e uma das principais componentes na viabilidade a longo prazo da marca (Krishnamurthi e Raj, 1991). A vantagem desta abordagem é que permite examinar os clientes para tentar identificar eventuais características comuns, “assim sendo o produto pode ser direccionado para os clientes potencialmente fieis” (Yelkur e Da Costa, 2001);

- h) *imagem* (Evans, 1959; Sirgy, 1982; Leisen, 2001) – Esta abordagem trata de dividir o mercado tendo em conta as “associações afectivas relacionadas com a imagem de marca” (Hanlan *et al.*, 2006). Esta área envolve a “*self-image*” ou “*self-concept*” do consumidor e a sua relação com a imagem do produto (Beane e Ennis, 1987). Segundo Beane e Ennis (1987), esta abordagem é uma combinação dos aspectos psicográficos e comportamentais com o consumidor;
- i) *situação* (Dickson, 1982; Gehrt e Shim, 2003) – As investigações sobre a *situação* no que diz respeito ao comportamento do consumidor são consideráveis (Cestre e Darmon, 1998). De acordo com Punj e Stewart (1983), as variáveis situacionais são de importância considerável para a segmentação (*research*). Em muitos dos casos, as variáveis situacionais foram confirmadas como determinantes da preferência e do comportamento de compra (Dickson, 1982);
- j) e *estilo social* (Cathelat e Wyss, 1989) – O *estilo social* considera que o comportamento de um indivíduo resulta de um comportamento regido por dois grandes factores: *i)* estrutura psicológica e mental; e *ii)* influência do seu ambiente. O contributo de um estudo dos estilos sociais na análise de *marketing* é uma “*inteligência funcional*” que tenta compreender “o que traz o produto, a marca ou informação para a vida das pessoas”. De acordo com Cathelat (1994) esta função varia de um perfil tipo a outro, criando, assim, uma segmentação útil do mercado. No entanto, os críticos salientam o facto de que esta abordagem requer um grande esforço intelectual devido ao seu carácter de síntese.

Convém referir que, em relação ao método escolhido, não se está perante “conceitos” fechados ou inalteráveis, pois podem ser acrescentadas novas combinações e variações que enriquecem continuamente aquele que é um dos aspectos mais importantes do *marketing*, e do qual depende, em grande medida, o sucesso ou o fracasso do processo da segmentação de clientes.

2.3.2. Segunda Fase: o Targeting

Esta fase consiste na escolha de alvos, ou seja, dos segmentos possíveis escolhe-se aqueles que mais interessam. Ela pode ser dividida nas seguintes etapas: 1) avaliar a atractividade de cada segmento; e 2) seleccionar os segmentos alvos.

Uma vez identificados os segmentos, é possível proceder à avaliação da atractividade de cada segmento. A atractividade do segmento é uma medida do potencial de um segmento “para produzir crescimento nas vendas e nos lucros” (McDonald e Dunbar, 2004).

Na avaliação da atractividade de diferentes segmentos, é necessário analisar se o potencial segmento possui as características que o tornam em geral atractivo, tais como o tamanho, o crescimento, a rentabilidade, a economia de escalas e o baixo risco.

Frank *et al.* (1972) e Wedel e Kamakura (1999) sugerem que o valor potencial de uma segmentação, para uma empresa, assenta nos seguintes critérios: *i)* identificável (*identifiability*): capacidade de reconhecer grupos distintos de consumidores no mercado, usando uma base de segmentação específica (*e.g.*, dimensão, poder de compra, etc.); *ii)* substancial (*substantiality*): capacidade de identificar uma parcela “suficientemente” grande e rentável, para garantir que vale a pena apostar nesse segmento; note-se que a dimensão “suficientemente” irá depender da estrutura e objectivos da empresa em causa; o maior segmento passível de explorar com o mesmo programa de *marketing*; *iii)* acessível (*accessibility*): nível ao qual é possível alcançar os segmentos alvos (fácil aproximação com o *marketing mix* definido); por outras palavras, podem ser efectivamente cobertos e servidos; *iv)* resposta (*responsiveness*): conseguida quando o segmento responde especificamente ao apelo que lhe é feito, *i.e.*, à campanha de *marketing* que lhes é exclusivamente direccionada; fundamental para a efectividade da estratégia, porque uma campanha só será efectiva se cada segmento é homogéneo e único na resposta a essa campanha; *v)* accionável (*actionability*): “medida” da atractividade de um segmento, isto é, o nível ao qual as necessidades e expectativas do segmento são consistentes com os objectivos e competências da empresa; e *vi)* estável (*stability*): um segmento tem de permanecer estável durante o tempo necessário para o identificar, desenvolver e implementar uma estratégia de *marketing* própria, e ainda para que essa estratégia dê resultados.

Estes critérios são frequentemente apontados para avaliar a qualidade e a pertinência dos segmentos, sendo mesmo considerados os determinantes da efectividade e rentabilidade da segmentação (Frank *et al.*, 1972; Dibb e Simkin, 1991; Hooley *et al.*, 1998; Wedel e Kamakura, 1999; Kotler, 2003).

De acordo com Wind e Cardoza (1974), os critérios sobre a avaliação dos segmentos são importantes, mas as pesquisas realizadas (*e.g.* Weinstein, 2004) demonstram que muitos negócios baseiam a sua estratégia de segmentação mais na intuição do que no *sound marketing planning* – que fornece uma forte base para aumentar as vendas e melhorar o desempenho global de *marketing* (Weinstein, 2004). De facto, tratar esses seis requisitos torna esta tarefa um desafio difícil que é referido pelos eruditos como a “instabilidade do segmento”⁷ (Wedel e Kamakura, 1999; Steenkamp e Hofstede, 2002; Blocker e Flint, 2007). De acordo com vários investigadores, os factores chave dessa “instabilidade” (ou “turbulência dos mercados”) deve-se ao facto de que os Mercados se estão a tornar cada vez mais dinâmicos (Eisenhardt e Martin, 2000; Douglas, 2001; Barnett e McKendrick, 2004; Voelpel *et al.*, 2005), caracterizados pela “mudança das necessidades dos clientes” e por “uma rápida evolução das preferências” (D'Aveni, 1995; Achrol e Kotler, 1999; Achrol e Etzel, 2003; Joshi e Campbell, 2003).

Independentemente da multiplicidade de critérios possíveis, o importante é que cada empresa segmente o seu mercado de acordo com aqueles que lhe pareçam relevantes (Brito, 1998).

Uma vez segmentado o mercado, a empresa deverá seleccionar os segmentos em que pretende actuar (*i.e.*, seleccionar os segmentos alvos). Wind e Thomas (1994) salientam que a selecção de segmentos é uma decisão crítica na gestão, porque “todas as outras componentes de uma estratégia de *marketing* vão segui-la”.

De acordo com Kotler e Dubois (2000), a definição do *Targeting* (escolha de alvos) está directamente ligado com o conceito de segmento. Pelo facto de observar um ambiente em rápida evolução com consumidores mais complexos e difíceis de classificar em segmentos precisos. Assim sendo, o gestor num processo de *Targeting* deve: *i)* conhecer os diversos cenários de homogeneidade dos segmentos; *ii)* conhecer

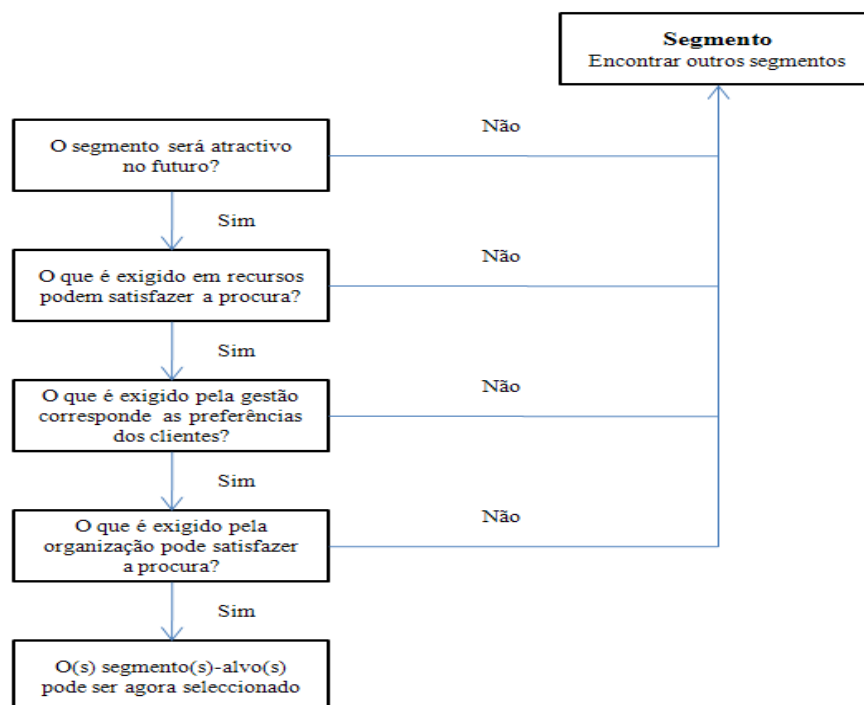
⁷ A instabilidade do Segmento refere-se a um estado de mudança nas necessidades dos clientes, bem como alterações na composição do segmento, reflectido pelas mudanças nos conteúdos do segmento e estrutura do segmento (consultar o artigo de Blocker e Flint (2007) para mais detalhes sobre este tema).

os tamanhos relativos dos diferentes segmentos, e isso no fim de *iii*) levar a um compromisso entre o tamanho do mercado representado por um ou vários segmentos e o grau de homogeneidade desejado para cada um desses segmentos-alvos.

A selecção de segmentos depende de uma multiplicidade de factores, deve-se, no entanto, realçar que, pela sua relevância, assumem um papel crucial nesta fase as aptidões e recursos da empresa e as características dos vários segmentos (Kotler *et al.*, 2005).

Por exemplo, no caso do mercado de *B2B*, Freytag e Clarke (2001) sugerem o seguinte modelo para seleccionar os segmentos (ver figura 2.3.). De acordo com os autores, este modelo, permite-nos ver o que uma empresa tem de percorrer para localizar e seleccionar os segmentos que melhor atendam às suas capacidades.

Figura 2.3. – Processo de Selecção de Segmentos no mercado B2B



Adaptado de Freytag e Clarke (2001)

2.3.3. Terceira Fase: O Posicionamento

Depois da selecção de segmentos, segue-se a terceira fase que consiste em desenvolver o posicionamento para cada segmento, ou seja, trata-se, em primeiro lugar,

de identificar, para cada segmento alvo, conceitos de posicionamento possíveis. E, em segundo lugar, de seleccionar, desenvolver e comunicar o conceito de posicionamento escolhido.

De acordo com Kotler (2003), o posicionamento é o acto de desenhar um conjunto de diferenças que distingam a empresa das suas concorrentes. Nessa ordem de ideias, Bannon (2003) propõe cinco factores chaves para que o posicionamento seja bem sucedido face à concorrência: *i)* clareza na ideia de posicionamento: clareza na compreensão da posição nas mentes dos clientes; *ii)* consistência da posição: é importante para os clientes saberem “qual a sua posição perante a empresa” (Bannon, 2003), sendo, por isso, necessária uma abordagem coerente e sustentável a longo prazo; *iii)* credibilidade do posicionamento: por exemplo, se a empresa pretende oferecer produtos de alta qualidade, os clientes “devem concordar que a qualidade dos produtos é realmente alta” (*idem*) , ou seja, a percepção por parte dos clientes é importante; *iv)* competitividade: os produtos devem (se é para alcançar uma vantagem competitiva) oferecer valor que os produtos concorrentes não oferecem; e *v)* comunicabilidade: a mensagem sobre a posição deve ser capaz de ser comunicada ao público-alvo. Como tal, a mensagem deve ser simples, atractiva e facilmente perceptível pelo público-alvo.

Keegan e Green (1999) salientam que o posicionamento é o processo de estabelecer a imagem de um produto na mente dos consumidores, relativamente à imagem da oferta dos concorrentes. Por outras palavras, o posicionamento é o espaço que um produto ocupa na mente do consumidor num determinado mercado. Relativamente ao posicionamento do produto para cada segmento, Cobra (1992) sugere os seguintes passos: *i)* identificar necessidades não satisfeitas e os agrupamentos homogéneos de consumidores com essas necessidades; *ii)* avaliar o potencial de compra desses grupos; *iii)* escolher os que se deseja atingir; e *iv)* identificar o posicionamento de cada produto concorrente e desenvolver uma estratégia capaz de diferenciar o produto nos seus respectivos segmentos de mercado.

Concretamente, recomenda-se que cada empresa se tente posicionar em segmentos que sejam simultaneamente atractivos e para os quais possua vantagens competitivas (Kotler *et al.*, 2005).

2.3.4. Quarta fase: O Desenvolvimento de um Programa de *Marketing (marketing mix)*

Esta fase consiste no desenvolvimento e implementação (Bannon, 2003) de um programa de *Marketing (marketing mix)*, ou seja, do conjunto de ferramentas de *Marketing* que cada empresa dispõe para prosseguir os seus objectivos dentro de cada um dos segmentos-alvo.

O termo *marketing mix* tornou-se popular com o artigo de Neil H. Borden – *The Concept of the Marketing Mix* publicado em 1964. Sendo depois revisto em 1975, por E. Jerome McCarthy, que reagrupou os “ingredientes” em quatro categorias que hoje são conhecidas pelos **4Ps**: **P**roduto/Serviço, **P**reço, **D**istribuição (“***Place***”) e **P**romoção (ver figura 2.4.), sendo que a interligação e interdependência destas quatro variáveis conduzem ao conceito de *marketing mix*. Esta componente da Estratégia de *Marketing* é geralmente definida no Programa de *Marketing* (Doyle, 2000). Assim, as políticas a definir são: Política de Produto – na política de produto dever-se-á definir as características intrínsecas do produto, descrever a embalagem, definir a marca e os serviços de pós-venda; Política de Preços – a política de preços engloba a definição do preço base e as condições de pagamento a praticar; Política de Comunicação (promoção) – na área de comunicação dever-se-á definir a estratégia de *media* e meios promocionais a usar; Política de Distribuição – a estratégia de organização da equipa de venda e a selecção dos canais e pontos de distribuição do produto (Kotler *et al.*, 2005).

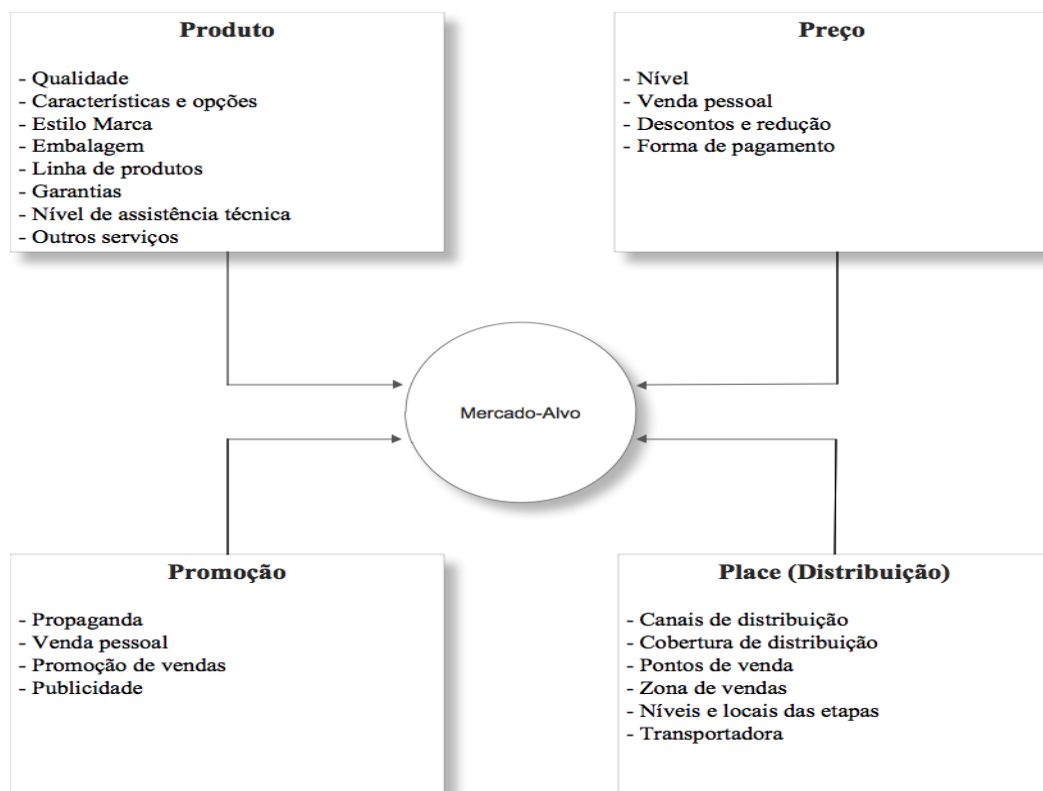


Figura 2.4. – O Marketing Mix (Adaptado de Baker, 2003)

A legitimidade do modelo dos 4Ps é posto em causa no caso dos Serviços por ser considerado insuficiente. É por isso que autores como Eigler e Langeard (1987) ou ainda Lovelock (1996) desenvolveram um novo modelo. Esse novo modelo tem em conta as especificidades dos *Servuction*⁸ (criação de um serviço) que são a intangibilidade, mas também a heterogeneidade e o seu carácter perfeitamente perecível.

O modelo dos 7Ps proposto por Booms e Bitner (1981) acrescenta às quatro categorias habituais, o **Processo** (*Process*): caracterizado pela interação com os clientes (por exemplo, atendimento, aconselhamento, horários de abertura, etc.); as **Pessoas** (*People*): capacidades da Força de Vendas (FV) (por exemplo, apresentação, formação, etc.); Suporte físico (*Physical evidence* ou *Physical support*): componentes materiais da loja (por exemplo, montra, organização das prateleiras, etc.), do serviço (por exemplo, relatório anual para um perito de contabilidade; extracto de conta, caderneta de cheques, ou cartão bancário para um banco), ou da identificação do pessoal, que faz parte integrante da produção para um serviço (por exemplo, o uniforme, etc.).

⁸ O *Servuction* é o sistema “necessário” a uma empresa para fabricar o serviço que ela própria coloca no mercado (Langeard e Eigler, 1999).

No entanto, alguns autores (*e.g.* Rafiq e Ahmed, 1995 *apud* Goi, 2009) criticaram o contributo conceptual dos 3Ps na medida em que as ideias que representam podem ser incluídas nos 4Ps originais: o processo seria essencialmente um problema ligado ao produto, enquanto que as pessoas seriam essencialmente ligadas à produção, ou seja, ao produto, ou, por vezes, à promoção, e, finalmente, o *Physical evidence* seria mais ao menos assimilado à promoção.

Para mais detalhes sobre o *marketing mix* e da problemática em relação aos 4Ps, recomendamos, por exemplo, o artigo de Goi (2009).

2.4. Áreas de Aplicação da Segmentação

A segmentação é um tema com uma grande diversidade de aplicação em *marketing*/vendas. No entanto, o grande número de referências sobre a segmentação torna “virtualmente impossível uma revisão de todas as aplicações até a data” (Wedel e Kamakura, 1999). Por isso, apenas vamos abordar algumas das áreas de aplicação que consideramos serem interessantes, quer do ponto de visto académico, quer do ponto de vista empresarial, sob o nosso ponto de vista. As áreas de aplicação abordadas nesta secção são as seguintes:

- a) Instituições financeiras e bancárias: em termos de investigações existem vários estudos realizados sobre segmentação nas áreas financeiras e bancárias, porque a segmentação permite, a título de exemplo, conhecer melhor o tipo da relação banco-cliente, de constituir portefólios de clientes confiados à força de vendas e de activar o dispositivo de distribuição (este tema será discutido com mais detalhes na secção 2.4.1);
- b) Indústria agro-alimentar: podemos destacar a título de exemplo algumas investigações sobre segmentação nesta área tais como: Segmentação da oferta de produtos frescos (Codron *et al.*, 2003); Aplicação da segmentação ao mercado de vinho: vinho verde (Brochado, 2002) de acordo com as denominações de origem (Rodriguez *et al.*, 2006);
- c) Automóvel: respeitante aos estudos e investigações empíricas realizadas na área automóvel coexistem duas abordagens da segmentação (Prieto, 2005): a

segmentação geográfica (Verboven, 1996; Ginsburgh e Weber, 2002) e a segmentação por produtos (Chanaron e Lung, 1995; Verboven, 2002). De acordo com Prieto (2005), é importante referir que os trabalhos e estudos disponíveis sobre a segmentação geográfica, considera os mercados nacionais, como os mercados geográficos, relevantes. A definição do mercado geográfico relevante baseia-se nos diferenciais de preços entre os países e as diversas barreiras à entrada (sistemas diferentes de homologação, necessidade de uma matrícula nacional, custos de administração e de transporte, sistema de distribuição específico). No que diz respeito à segmentação por produto, no caso do mercado europeu, por exemplo, os segmentos “clássicos” são os segmentos: *i)* inferiores, *ii)* médios inferiores, *iii)* médios superiores, *iv)* superiores, *v)* desportivo / coupé e *vi)* monoespaço / camionetas (Verboven, 2002);

- d) Internet: existe uma grande variedade de estudos realizados sobre a segmentação na Internet. Podemos destacar, a título de exemplo, os seguintes trabalhos de investigação: Segmentação de utilizadores em função da frequência de uso, da proporção de cliques que se destinam às diferentes classes de conteúdos e aos diversos tipos de artigos, e do padrão de visita das suas sessões (Rebelo, 2006); agrupar secções de usuários Web para prever os comportamentos futuros do utilizador (Mobasher *et al.*, 2000); segmentação dos consumidores através das compras *on-line* (Ruiz e Lassala, 2006); tipologia de clientes do ciberespaço (Barnes *et al.*, 2007), entre outros;
- e) Indústria farmacêutica: de acordo com vários autores, a indústria farmacêutica está sujeita a vários desafios: expiração das patentes, a pressão sobre os preços, concorrência dos genéricos, citando apenas alguns exemplos, no qual o papel da segmentação de mercado teve um papel importante contribuindo, por exemplo, para aumentar a eficiência dos canais de distribuição e responder melhor às necessidades dos clientes. Em termos de investigações, podemos destacar os seguintes: Segmentação de mercado farmacêutica na idade da Internet (Lerer, 2002);
- f) Turismo: na indústria do turismo, existem vários estudos realizados sobre a segmentação (Legohérrel e Wong, 2006). Como consequência da diversidade da actividade turística e dos seus participantes (visitantes), a segmentação de

mercado constitui uma das principais componentes da investigação em turismo (Mo *et al.*, 1994). Em termos de investigação académica, a nível nacional sobre o turismo, podemos, por exemplo, destacar investigações realizadas de acordo com os fluxos turísticos (Águas, 2005), sendo os segmentos de mercado agrupados de acordo com a atractividade dos destinos; Segmentação do mercado turístico (Hanlan *et al.*, 2006); Segmentação do mercado turístico do património baseado no tipo de atracções visitadas (Kerstetter *et al.*, 1998), entre outros;

- g) Transportes: *ferroviários* Pas e Huber (1992) identificaram cinco tipos de viajantes que são: (1) o viajante funcional, (2) o excursionista do dia (*day tripper*), (3) os amantes dos comboios, (4) o lazer (hedónico) viajante, e (5) a família viajante; *aéreos* – companhias aéreas (Bunning *et al.*, 1985), segmentação de mercado no sector aeroportuário europeu (Freathy e O'Connell, 2000); *rodoviários* – transporte de frete (Matear e Gray, 1995), com critérios de segmentação bem definidos; pode ser uma importante ferramenta de apoio para a tomada de decisões, contribuindo com informações relevantes sobre os valores dos fretes pagos no mercado conforme as características da operação e dos diferentes serviços oferecidos pelos transportadores;
- h) Ramo têxtil: moda/retalho (Quinn *et al.*, 2007); moda e vestuário (Paço, 2002); neste artigo, a temática da segmentação de mercado é direccionada para investigar a sua aplicação no caso do vestuário de moda, ressaltando a importância da identificação de distintos segmentos de consumidores, tendo em conta as suas atitudes em relação à compra de vestuário.

2.4.1. Segmentação de Mercados na Banca

As instituições bancárias têm um papel primordial nas economias, realizando grande parte da intermediação de fundos para aqueles que têm necessidades líquidas de financiamento. Este papel dos bancos é especialmente importante para os agentes económicos com necessidade de financiamento (de despesas de consumo e

investimento), mas cujo acesso aos mercados dos títulos é limitado, como é o caso das famílias e das empresas de pequena e média dimensão (Banco de Portugal, 2003)⁹.

Esta área sempre teve um particular interesse pelos investigadores do *marketing* bancário¹⁰, sendo que durante os últimos 30 anos muitas investigações foram realizadas neste sector, focalizando-se nos clientes de bancos comerciais (Edris e Almahmeed, 1997). Uma variedade de estudos de *marketing* foi realizada sobre os clientes, incluindo atitude dos serviços bancários, fidelidade bancária, satisfação e *multiple bank users* e a segmentação de mercado (*e.g.* Laroche e Manning, 1984; Stanley *et al.*, 1985; Ruddick, 1986; Jain *et al.*, 1987; Yava, 1988; Denton e Chan, 1991; Kaynak *et al.*, 1991).

Num sector tão dinâmico como o sector bancário, a segmentação teve sempre um papel importante no desenvolvimento de novas ofertas de produtos. Aliás, na indústria bancária, como noutros serviços industriais, a segmentação é considerada como o melhor “caminho para operacionalizar o conceito de *marketing*” (Edris e Almahmeed, 1997) e fornece um guia para as estratégias de *marketing* bancário e alocação de recursos nos mercados e serviços (Wind, 1978; Dickson e Ginter, 1987; Rao e Wang, 1995). De acordo com Lindgreen e Antioco (2005), a segmentação foi a actividade chave que permitiu, por exemplo, ao *First European Bank*¹¹ perceber melhor os tipos de clientes com os quais estavam a tratar/negociar, e permitiu adaptar conformemente as suas comunicações de *marketing*. Os elementos mais importantes a ter em consideração são: a atitude e necessidade dos clientes; os aspectos sócio-demográficos; a actual e potencial rentabilidade; e os comportamentos em termos de uso dos canais de distribuição e produtos.

Após a revisão da literatura, constatamos que muitos estudos sobre a segmentação (individual) de clientes usavam abordagens comportamentais orientadas para a decisão (*e.g.* Stanley *et al.*, 1985; Laroche *et al.*, 1986; Ruddick, 1986; Yavas, 1988; Kaynak *et al.*, 1991). Da mesma forma, a abordagem comportamental e a abordagem direccionada para a decisão são muito úteis e produtivas para qualquer

⁹Fonte: Banco de Portugal (2003), “Inquérito aos Bancos sobre o Mercado de crédito”, in www.bportugal.pt/pt-PT/EstudosEconomicos/Publicacoes/IBMC/Documents/Inq_banc_apresent_p.pdf.

¹⁰ O *marketing* bancário é um *marketing* dos serviços, mesmo apresentando especificidades bem estabelecidas, empresta um bom número de ferramentas e de esquemas de análise de alcance mais gerais, ou seja, da teoria do *marketing* fundamental.

¹¹ De acordo com o Lindgreen e Antioco (2005), o “*First European Bank*” é parte de um grupo financeiro mundial que actua em mais de 60 países e tem mais de 100,000 colaboradores.

estratégia de segmentação de “um mercado *business-customer* de um banco” (Edris e Almahmeed, 1997).

O objectivo principal da segmentação nas instituições financeiras e bancárias é o de ajustar a organização comercial e o dispositivo da rede multicanal (*call-center*, *e-marketing*) aos segmentos e subsegmentos identificados previamente (Tassel, 2004).

Em geral, os bancos conhecem muito bem as informações dos seus clientes e trabalham desde há muito tempo em produzir segmentações finas dos seus clientes, visando propor a cada um “dos segmentos” (ou “tipos” de clientes) serviços que, por um lado, os satisfarão, por outro lado, serão rentáveis para a organização. Ou seja, uma segmentação de clientes adequada e eficaz pode proporcionar uma diferença significativa num mercado (financeiro) cada vez mais competitivo. De acordo com vários especialistas na área, um factor chave é conseguir encontrar “como pode diferenciar-se dos seus concorrentes”

A “segmentação bancária” permite por exemplo: *i*) determinar diferentes perfis de clientes (Seggern e Hadaway, 2004), permitindo a catalogação dos clientes em determinados segmentos: de muito rentável a muito pouco rentável, por exemplo; *ii*) focalizar melhor as suas ofertas e entrega aos clientes (*Harvard Business Essentials*, 2006); *iii*) servir os seus clientes de modo mais eficiente¹²; *iv*) perceber e responder melhor as necessidades dos seus clientes: o facto de conhecer as necessidades dos clientes torna possível a descoberta de novas oportunidades; *v*) efectuar uma pesquisa mais eficiente da rentabilidade dos clientes; *vi*) melhorar a gestão da rede multicanal: a segmentação de clientes permitirá ao profissional de *marketing* definir melhor tanto o canal de distribuição como o canal de comunicação preferida pelos diferentes segmentos de clientes (Tassel, 2004); *vii*) minimizar o *score*¹³ (ou grau) de abandono de clientes (*Customer attrition*¹⁴): tratar-se da detecção dos segmentos de clientes que manifestam um desinteresse com o banco; *viii*) melhorar a fidelização de clientes; e *ix*) estabelecer análises prospectivas das tendências em necessidades de financiamento e das preferências dos clientes.

¹² Fonte: www.santandertotta.pt.

¹³ Um *Score* (pontuação) é uma maneira de expressar os resultados de um modelo num único número. Os *scores* podem ser usados para classificar uma lista de clientes.

¹⁴ O *Customer attrition* (também conhecido como *customer churn*, *customer turnover*, ou *customer defection*): é um termo inglês usado nos negócios para descrever a perda de clientes ou consumidores.

Para captar o interesse dos clientes e “melhor os fidelizar”, os departamentos de *marketing* utilizam modelos de segmentação cada vez mais sofisticados (Tassel, 2004), usando, por exemplo, técnicas de *Data Mining*, conforme discutido no próximo capítulo.

A maioria dos grandes bancos apresenta uma “conhecida” segmentação que diferencia os seus clientes. Basicamente, existem os segmentos de “*Consumer*” – Particulares, “*Private*” e “*Corporate*” – empresas, apesar de existirem outras subdivisões, que variam com cada banco.

O segmento de *Consumer* trabalha com o grande público. Em relação ao segmento *Private Banking*, é um serviço especial de gestão de património oferecido pelos bancos, incluindo atendimento personalizado e possibilidade de investimentos em aplicações financeiras mais sofisticadas e “rentáveis”. Os clientes que compõem este segmento são principalmente empresários e profissionais liberais. O *Corporate* é um segmento composto por empresas que recebem dos bancos diversos serviços especializados; algumas instituições financeiras subdividem este segmento de acordo com “a facturação” das empresas-clientes. Os produtos e serviços oferecidos aos clientes deste segmento são muito diferentes do que os elaborados para os particulares.

Em geral, o papel do banco neste caso é mais o de acompanhar, prestar consultoria e serviços do que na venda a retalho, por exemplo, cujo maior interesse é a venda de produtos.

CAPÍTULO III

*“Who (ever) has information fastest
and uses it wins.”*

(Don McKeough)

Data Mining para Segmentação

O rápido crescimento de grandes quantidades de dados recolhidos e armazenados em grandes e numerosas bases de dados, acabou por superar a nossa capacidade de compreensão humana (Han e Kamber, 2006). Devido à esta vasta quantidade de dados e da iminente necessidade de transformar esses dados em informações e conhecimentos úteis, são necessários instrumentos válidos para a sua modelação e análise (Giudici, 2003). Esta abundância de dados reflecte-se por exemplo em problemas relacionado com a segmentação.

De acordo com vários autores (*e.g.* Fayyad *et al.*, 1996), para lidar e tratar tais dados, o **Data Mining** (DM) é a ferramenta mais apropriada para extrair informação e conhecimento útil desses dados. O *Data Mining* (DM) é o processo de análise de dados para obtenção de conhecimento, também conhecido por Extracção de Conhecimento de Dados (ECD).

O DM é usado por uma grande variedade de indústrias e sectores de diversas áreas, como por exemplo as telecomunicações, a medicina, as finanças, a área farmacêutica e o marketing (Fayyad *et al.*, 1996). O DM tem sido utilizado na abordagem de todos os aspectos de marketing, vendas, e gestão de produtos (Thelen *et al.*, 2004), sendo o marketing uma das áreas de maior impacto do DM.

Para efeitos deste presente trabalho, assumimos que o objectivo do *Data Mining* é de permitir às empresas “melhorar” os seus processos de marketing, vendas e suporte às operações de clientes através de uma melhor compreensão dos seus clientes (Berry e Linoff, 2004). Compreender os clientes envolve recolher dados sobre as interacções

entre os clientes e a empresa, analisar esses dados para transformá-los em conhecimento e, finalmente, aplicá-lo nas actividades de *marketing* (Böttcher *et al.*, 2009).

Uma das aplicações mais comuns de DM em *marketing* é a segmentação de clientes (Thelen *et al.*, 2004). Para esse fim, são usados métodos de *clustering* (agrupamento) discutido no próximo capítulo, cuja função é dividir um conjunto de dados em subconjuntos homogéneos. Em problemas de segmentação, os dados representam características dos clientes.

De seguida, discutir-se-á em mais pormenor o conceito de *Data Mining* (Secção 3.1). A seguir, descrevemos as aplicações do *Data Mining* (Secção 3.2), sendo também abordado as principais metodologias usadas na resolução de um problema de *Data Mining* para a segmentação (Secção 3.3).

3.1. Conceito de Data Mining

Podemos definir o *Data Mining* como um conjunto de técnicas e ferramentas aplicadas ao processo de extracção de conhecimento implícito, previamente desconhecido, de uma base de dados¹⁵ com vista a utiliza-lo, por exemplo em decisões empresárias, e a maximizar o valor dos dados que são recolhidos (Piatetski-Shapiro, 1991). O DM consiste num processo analítico de exploração¹⁶ e análise de grandes quantidades de dados, por meios automáticos e semi-automáticos, na busca de padrões (Fayyad *et al.*, 1996) e regras consistentes que possam ser importantes para a resolução de um determinado problema.

Na literatura, o termo *Data Mining* confunde-se frequentemente com o termo **Descoberta de Conhecimentos em Bases de Dados (DCBD/KDD¹⁷)**, pelo facto de o DM se ter tornado num conceito cada vez mais popular no meio industrial e empresarial. O DCBD é uma área multidisciplinar que incorpora técnicas utilizadas em diversas áreas como Base de Dados, Inteligência Artificial (IA) e Estatística (Degroot, 1986). Esta área tem despertado um crescente interesse devido ao aumento exponencial de dados

¹⁵ A *base de dados* é entendida normalmente como o suporte físico da informação armazenada (num computador).

¹⁶ www.statsoft.com.

¹⁷ Knowledge Discovery in Database (KDD) é traduzido em português por *Descoberta de Conhecimentos em Bases de Dados* (DCBD) ou por *Extracção de Conhecimento de Dados* (ECD).

armazenados e à necessidade económica e científica de extrair informação. Os métodos estatísticos tradicionais não têm a capacidade e escala para analisar esses dados e, portanto, foi necessário desenvolver as metodologias e ferramentas modernas.

O DCBD é constituído por várias etapas (como ilustrado na figura 3.1.) que são executadas de forma interactiva e iterativa (Fayyad *et al.*, 1996). De acordo com Brachman e Anand (1996) as etapas são interactivas porque envolvem a cooperação do responsável pela análise de dados, cujo conhecimento sobre o domínio orientará a execução do processo. Por sua vez, a iteração deve-se frequentemente ao facto deste processo não ser executado de modo sequencial, pois envolve repetidas selecções de parâmetros, aplicações das técnicas de DM e posterior análise dos resultados obtidos, no fim de “refinar” os conhecimentos extraídos.

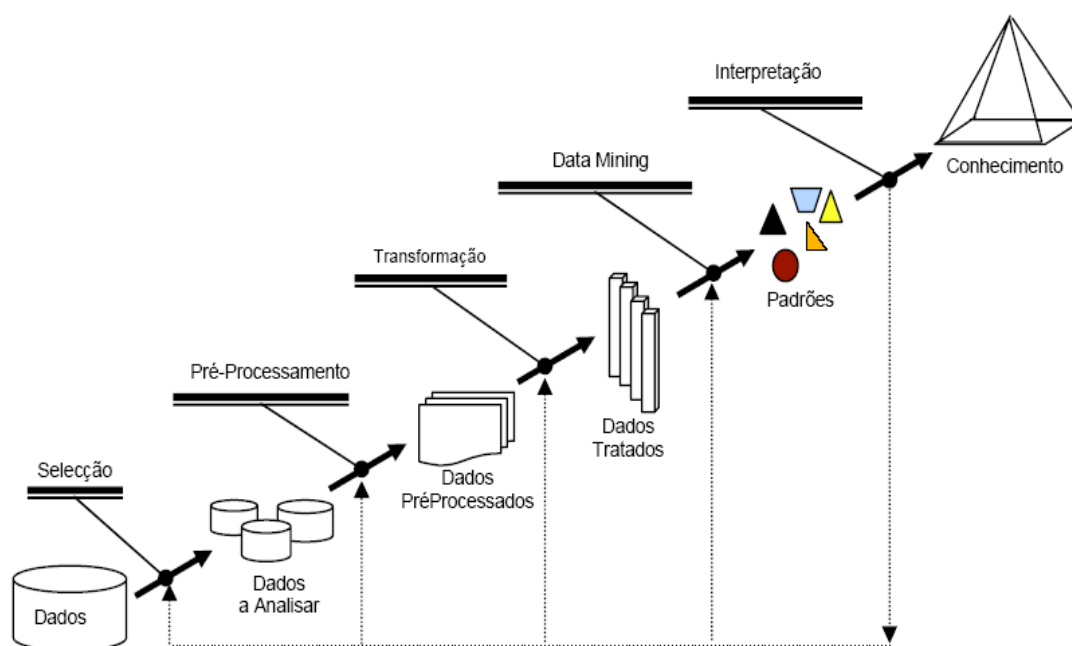


Figura 3.1. – Fases do processo de KDD/DCBD (adaptado de Fayyad *et al.*, 1996)

De acordo com Fayyad *et al.*, (1996), o *Data Mining* é um passo do processo de KDD. Actualmente, no entanto, utiliza-se DM como sinónimo de KDD.

Geralmente, quando se trabalha com DM, os dados estão organizados em tabelas contendo linhas e colunas em formato matricial. Normalmente, o DM distingue-se das outras técnicas de análise de dados na forma como explora as relações entre os dados. Nas ferramentas tradicionais de análise disponíveis, o utilizador constrói hipóteses sobre relações específicas e então corrobora-as ou refuta-as através dos resultados da

ferramenta utilizada. De acordo com Rygielski *et al.* (2002), o processo de DM é responsável pela geração de hipóteses, o que potencia maior rapidez, aperfeiçoamento, autonomia e fiabilidade aos resultados.

Dependendo do tipo de padrão e do objectivo da aplicação, os problemas de DM são abordados usando uma (ou mais) tarefas diferentes (Han e Kamber, 2006). No *Data Mining* uma tarefa corresponde num tipo de problema de descoberta de conhecimento para o qual se pretende arranjar uma solução.

Em geral, as tarefas de DM podem ser classificadas em duas categorias, dependendo do objectivo: *preditivas* (envolve usar algumas variáveis para prever valores desconhecidos ou futuros de outras variáveis) e *descritivas* (foca-se em encontrar padrões interpretáveis que descrevam dados que podem ser interpretados por seres humanos) (Kantardzic, 2003; Han e Kamber, 2006):

A) Métodos Preditivos:

- Classificação
- Regressão
- Análise de séries temporais

B) Métodos Descritivos:

- Clustering (será discutido no próximo Capítulo)
- Associação (*Association Rule Discovery*)
- Sequência (*Sequential Pattern Discovery*)
- Sumarização

Tabela 3.1. – Tarefas de *Data Mining*

Tarefas	Descrição
Classificação	A <i>classificação</i> é o processo de extrair de dados um modelo (ou função) que descreve e distingue classes de objectos e que será usado para prever a classe de novos objectos (Han e Kamber, 2006).
Regressão	A <i>regressão</i> é uma metodologia estatística que é frequentemente usado para a previsão, embora existam outros métodos também (Han e Kamber, 2006).
Análise de séries temporais	A <i>análise de séries temporais</i> consiste em identificar regularidades e características temporais interessantes que estão escondidas nos dados.
Clustering	O <i>clustering</i> é o processo de agrupar os dados em classes ou clusters de modo que

(agrupamento)	os objectos dentro do cluster tenham uma alta similaridade em comparação com um outro e, no entanto que sejam diferentes dos objectos dos outros Han e Kamber, 2006). Ou seja, a tarefa de <i>clustering</i> consiste no agrupamento automático de dados em subconjuntos de objectos com características bastante semelhantes.
Associação	A tarefa de <i>associação</i> consiste em identificar regras que descrevem afinidades entre as combinações de elementos de dados (Ye, 2003). A tarefa de <i>associação</i> é frequentemente usado em aplicações como a análise do Cabaz de Compras (<i>Market Basket Analysis</i>), para medir as associações entre produtos adquiridos por um consumidor particular, e ainda para a análise da sequência de cliques de página <i>web</i> (<i>Web clickstream analysis</i>), para medir as associações entre páginas visitadas sequencialmente por um visitante de sítio Web (Giudici, 2003).
Sequência	A tarefa da <i>sequência</i> de padrões é uma regra de associação que utiliza o factor tempo. É usado para encontrar relações temporais nos dados (Rud, 2001)
Sumarização	A tarefa de <i>sumarização</i> envolve métodos para encontrar uma descrição compacta para um subconjunto de dados (Fayyad <i>et al.</i> , 1996). Ou seja, a tarefa de <i>Sumarização</i> consiste, em geral, em métodos estatísticos mais simples (<i>e.g.</i> Média, Desvio, Correlação) que trata de descrever as características de um grupo ou conjunto de dados (<i>e.g.</i> Distribuição dos Assinantes da Revista “X” por Regiões).

Convém referir que a lista de tarefas, apresentadas na tabela 3.1., não é fechada, pois, dependendo do autor, é possível encontrar novas tarefas de DM na literatura.

As técnicas de DM surgem como *algoritmos* ou metodologias que são aplicadas na implementação das tarefas referidas anteriormente. Existem vários algoritmos possíveis para cada uma dessas tarefas, sendo que a escolha de um determinado algoritmo pode depender por exemplo do tipo dos dados disponíveis, do propósito específico da aplicação, entre outros factores.

Um *algoritmo* consiste numa lista ou sequência de instruções que podem ser usadas para realizar uma tarefa. Deve realçar-se que não existe um algoritmo universal de *Data Mining* que resolva, de forma eficiente, todos os problemas (Harrison, 1998). A escolha de um determinado algoritmo é de certa forma uma arte (Fayyad *et al.*, 1996), uma vez que existem diferentes algoritmos para as mesmas tarefas de DM com vantagens e desvantagens intrínsecas.

As vantagens na utilização de DM são: *i)* permitir analisar grandes bases de dados (Berry e Linoff, 2004); *ii)* apresentar um método automático de descobrir padrões nos dados (Kantardzic, 2003); *iii)* permitir a criação de modelos precisos (isso tendo em

conta a qualidade dos dados) (Berry e Linoff, 2004); *iv*) os modelos¹⁸ são construídos e actualizados em tempo útil mesmo quando tratando grandes quantidades de dados; *v*) produzir mais vantagens competitivas (pelo facto de fazer um melhor uso dos dados) (Rud, 2001); *vi*) permitir conhecer e identificar melhor os seus clientes (Berry e Linoff, 2004); *vii*) descobrir mercados inexplorados com bom potencial; e *viii*) maximizar o esforço comercial e adaptação mais “fina” às expectativas do mercado. As desvantagens mais referenciadas são: *i*) custos ligados às aplicações de Data Mining; *ii*) complexidade das ferramentas (Kantardzic, 2003); *iii*) o desafio da preparação dos dados; *iv*) interacção muito forte com analistas humanos; *v*) uso indevido da informação / informação imprecisa (McDonough, 2002); e *vi*) questões de privacidade e segurança (Rygielski et al., 2002).

A seguir, vamos abordar algumas aplicações de técnicas de *Data Mining*.

3.2. Aplicações de Técnicas de Data Mining

Muitas organizações já perceberam que o conhecimento armazenado na base de dados é fundamental para apoiar as várias decisões organizacionais. Em particular, o conhecimento sobre os clientes que pode ser obtido a partir desses dados é crítico para a função de *marketing* e vendas.

A seguir, são relacionadas as principais áreas de interesse na utilização de *Data Mining* (Mannila, 1997; Cratochvil, 1999; Dias, 2002):

- a) *Marketing* / Vendas: com as ferramentas de DM e de *Business Intelligence* (BI) tornou-se possível: *i*) descobrir e identificar padrões de comportamento de consumidores; *ii*) ajudar no desenvolvimento de novos produtos (Thelen et al., 2004); *iii*) associar comportamentos à características demográficas de consumidores; *iv*) reforçar a aquisição de clientes com o marketing directo; *v*) análise de vendas cruzadas (associações e correlações entre vendas de produtos); *vi*) estabelecer contactos multi-canal, para melhorar desenvolvimento do cliente

¹⁸ Um modelo é o output de um método de *Data Mining*, que em termos abstractos corresponde (o método) a um algoritmo. Esse método pretende resolver uma determinada tarefa/problema de Data Mining.

com o *cross-selling* e *up-selling*¹⁹ de produtos (Thelen *et al.*, 2004), bem como aumentar a retenção dos melhores clientes; e *vii*) identificar novas oportunidades de vender serviços adicionais;

- b) Detecção de fraudes: muitas fraudes “óbvias” podem ser encontradas sem DM como por exemplo na compensação de cheque por pessoas falecidas, mas a detecção de padrões mais subtis de fraude, como o caso do uso fraudulento de cartões de crédito (Kantardzic, 2003) pode ser mais dificilmente detectado;
- c) Medicina: o DM tem sido muito aplicado na área da medicina, e permitiu identificar por exemplo: *i*) terapias médicas de sucesso para diferentes doenças; *ii*) factores causa/efeito; *iii*) novas abordagens para os tratamentos (Ferreira *et al.*, 2006); *iv*) determinar padrões, como é o caso do DNA (*Deoxyribonucleic Acid*) e dos genes (Fayyad *et al.*, 1996); e *v*) caracterizar o comportamento do paciente para prever visitas, identificar terapias. Podemos também destacar, à titulo de exemplo, as investigações sobre a compreensão do fenómeno da dor recorrendo a técnicas de *Data Mining* (Ferreira *et al.*, 2006);
- d) Instituições financeiras e bancárias: de acordo com vários profissionais da área, o DM permite ajudar a: *i*) customização das diversas ofertas promocionais. Por exemplo, os emails directos podem ser personalizados de acordo com o segmento dos titulares da conta no banco; *ii*) melhorar as taxas de resposta das campanhas de emails; *iii*) reduzir o tempo necessário para classificar os clientes, este, por sua vez, vai permitir melhorar a eficiência da força de vendas do grupo-alvo; *iv*) ajudar na optimização do seu portfólio de serviços e canais de entrega; *v*) efectuar uma melhor segmentação e *targeting* de clientes ao analisar melhor os padrões de compra dos clientes ao longo do tempo; *vi*) detectar novas tendências de compras favorecendo atitudes pró-activas como por exemplo acrescentar valor aos produtos e serviços existentes e lançar novas ofertas de produtos e serviços; *vii*) detectar padrões de uso de cartão de crédito fraudulento; *viii*) identificar clientes “leais”, *ix*) determinar gastos com cartão de crédito por grupos de clientes; e *x*) encontrar correlações escondidas entre

¹⁹ Os modelos de *Cross-sell* são usados para prever a probabilidade ou valor de um actual cliente de comprar um produto ou serviço da mesma empresa (*cross-sell*). Enquanto que os modelos de *Up-sell* são usados para prever a probabilidade ou valor de um cliente para comprar mais do mesmo produto ou serviço (Rud, 2001).

diferentes indicadores financeiros (Dias, 2002).

De acordo com Rygielski *et al.*, (2002), os bancos podem utilizar os conhecimentos descobertos nas bases de dados em várias aplicações, incluindo: *i)* o *marketing* de cartões: ao identificar segmentos de clientes, emissores e adquirentes de cartões será possível melhorar a rentabilidade com programas mais eficiente de aquisição e retenção, desenvolvimento de produto-alvo, e preços customizados; *ii)* a avaliação e rentabilidade dos titulares dos cartões: “As entidades emitentes de cartões podem tirar partido da tecnologia de *Data Mining* para fixar o preço dos seus produtos de modo a maximizar os lucros e minimizar a perda de clientes”²⁰; *iii)* a detecção de fraudes: ao analisar as transacções passadas determinadas como fraudulentas, o banco pode identificar padrões que torna essa tarefa mais simples e eficaz; e *iv)* a previsão da Gestão de ciclo de vida: o *Data Mining* ajuda os bancos a prever o *Lifetime Value*²¹ (LTV) de cada cliente e servir cada segmento adequadamente (por exemplo, oferecendo descontos e promoções especiais);

- e) Apólice de seguro: na análise de reivindicações - determinar quais procedimentos médicos são reivindicados, prever quais clientes comprarão novas apólices, identificar padrões de comportamento de clientes perigosos, e identificar comportamento fraudulento (Viveros *et al.*, 1996);
- f) Industria de retalho: os dados no retalho “abrangem um amplo espectro” (incluído vendas, clientes, empregados, transportação de mercadorias, consumo, e serviços) (Han e Kamber, 2006). Segundo esses autores, existe um grande diversidade de utilização de técnicas de DM para a industria de retalho tais como: *i)* Análise da eficácia das campanhas de vendas; *ii)* Análise multidimensional de vendas, clientes, produtos, tempo e retenção de cliente por regiões; *iii)* análise de fidelização de clientes; e *iv)* recomendação de Produto e referências cruzadas de artigos (*cross-referencing of items*);
- g) e Web: existem muitas investigações direccionadas à aplicação de DM na *web*, tais como: (Mobasher *et al.*, 2000; Sarawagi e Nagaralu, 2000; Rebelo, 2006;

²⁰ Fonte: www.data-miners.com.

²¹ No Marketing, o *Lifetime Value* (LTV) de um cliente é o valor actual (geralmente expresso em moeda) dos lucros futuros que podem ser derivados de um cliente baseado nos lucros que foram recebidos desse cliente no passado.

Ruiz e Lassala, 2006).

Um inquérito realizado pelo *KDnuggets*²² em 2008 sobre as aplicações de DM revelam que as duas principais utilizações do *Data Mining* são relacionados com CRM - *Customer Relationship Management* e com a área bancária (ver na figura 3.2.).

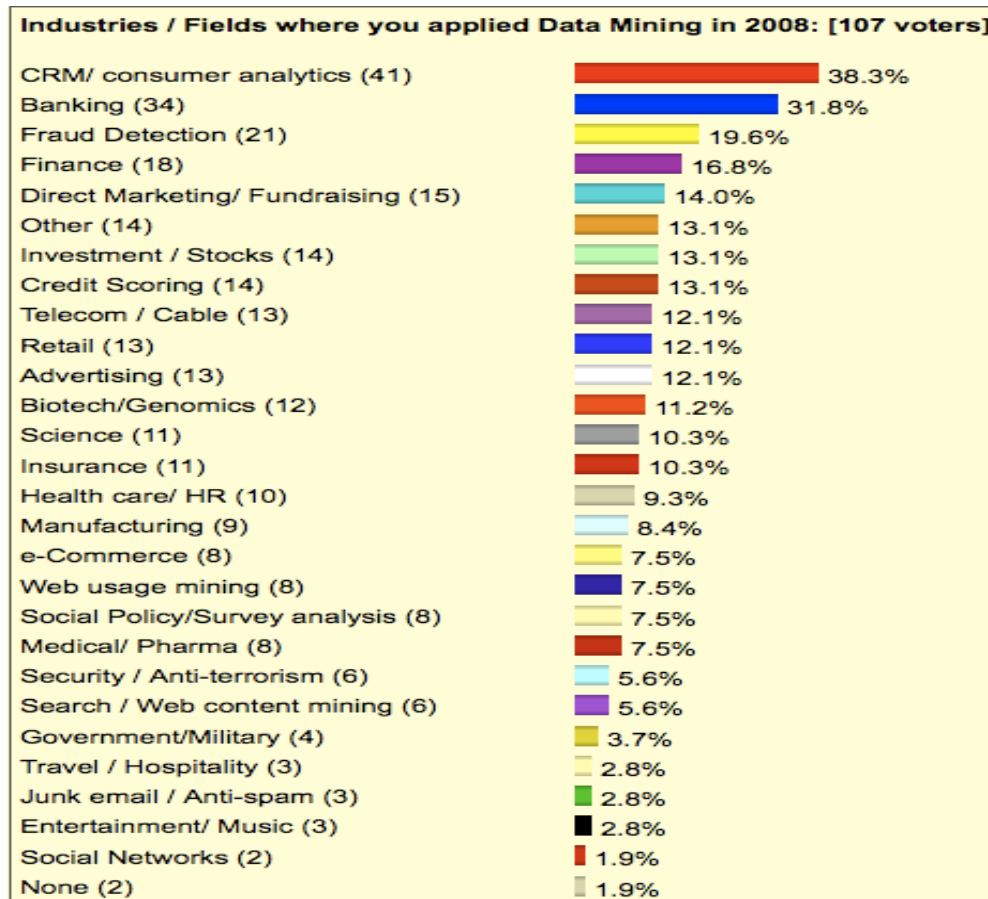


Figura 3.2. – Inquérito sobre as aplicações de *Data Mining* (por indústria) realizado no mês de Dezembro de 2008) pelo site *kdnuggets.com*

Ou seja, é interessante constatar que, em geral, as principais áreas de utilização do DM estão relacionadas com o *marketing* e a área comercial. Na nossa opinião, acreditamos que as ferramentas de DM vão continuar a evoluir e a desenvolver-se nestas áreas tão versáteis e complexas que são as vendas e o *marketing*.

Uma descrição mais detalhada das aplicações de DM pode ser encontrada em

²² www.kdnuggets.com.

Berry e Linoff (2004); Han e Kamber (2006).

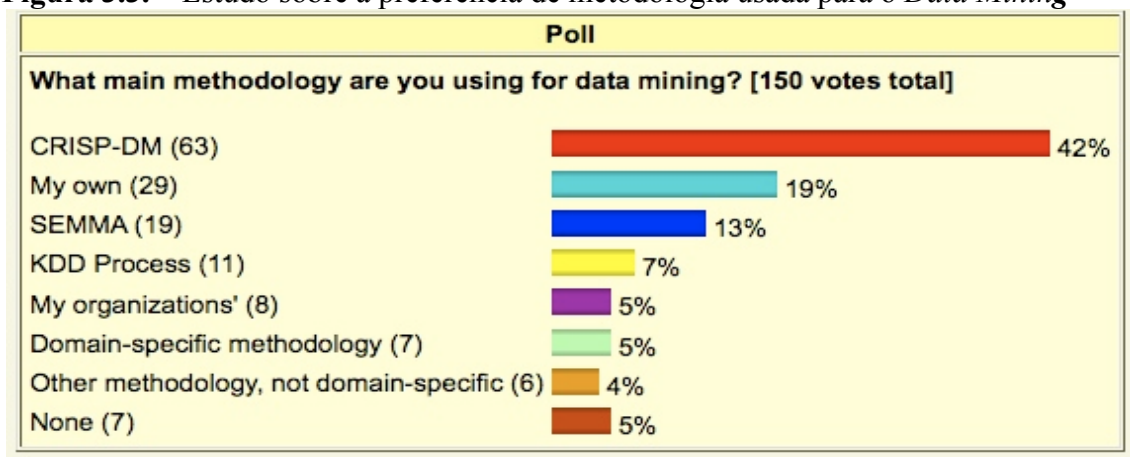
3.3. Metodologias para o Data Mining

O DM é um processo complexo que envolve várias tarefas, métodos e pessoas, e para ser bem sucedido em qualquer problema é crucial seguir uma metodologia bem definida a partida. Várias metodologias de *Data Mining* foram propostas para tentar moldar as etapas de desenvolvimento e implementação num processo estandardizado. Salientamos que o número de instrumentos e metodologias tem aumentado em número significativo, e vem responder de forma cada vez mais eficiente aos interesses dos empresários.

Actualmente encontram-se duas metodologias disseminadas e bem definidas para o desenvolvimento prático do DM: a metodologia CRISP-DM e a metodologia SEMMA.

Em termos de preferência e utilização de metodologia pelos profissionais encontra-se disponível um estudo realizado na Internet onde se pode verificar um domínio da metodologia CRISP-DM. No entanto, podemos salientar que o SEMMA está a ser cada vez mais utilizado pelo profissionais (ver figura 3.3.).

Figura 3.3. – Estudo sobre a preferência de metodologia usada para o *Data Mining*



Fonte: Estudo realizado por www.kdnuggets.com, 25 de Agosto - 10 de Setembro de 2007

3.3.1. CRISP-DM

A metodologia CRISP-DM (**C**Ross-**I**ndustry **S**tandard **P**rocess for **D**ata **M**ining) é adoptada por um consórcio internacional de empresas, incorporando conhecimento prático, de forma a responder aos requisitos dos utilizadores. Esta metodologia não centra os princípios unicamente na tecnologia, mas também nas necessidades dos utilizadores para a resolução de problemas de negócio (Han e Kamber, 2006).

O CRISP-DM define-se num conjunto de seis etapas/fases para o desenvolvimento estruturado e metodológico de projectos de *Data Mining* (ver figura 3.4.):

- a) o *Estudo do Negócio* – obter uma visão clara das necessidades a satisfazer;
- b) o *Estudo dos Dados* – determinar quais os dados disponíveis (e onde se encontram) para encontrar respostas;
- c) a *Preparação dos Dados* – Adaptar e formatar os dados de forma apropriada às respostas a encontrar;
- d) a *Modelação* – criar modelos explicativos das necessidades a satisfazer;
- e) a *Avaliação* – testar os resultados encontrados contra os objectivos do projecto;
- f) e a *Implementação (Distribuição dos Resultados)* – disponibilizar os resultados do projecto aos decisores.

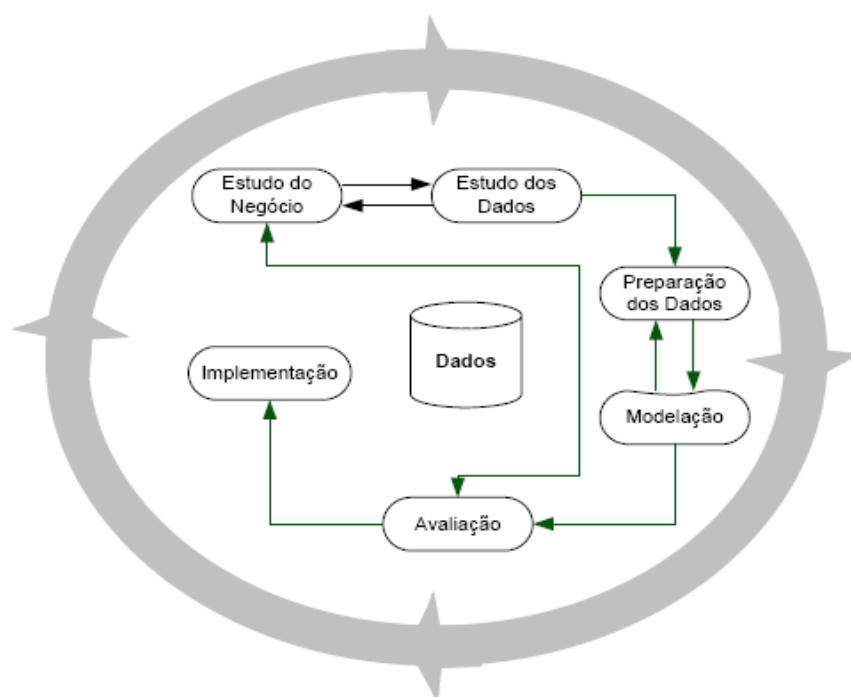


Figura 3.4. – Ciclo de vida da metodologia CRISP-DM (adaptado de Chapman *et al.*, 2000)

Para mais detalhes sobre a metodologia CRISP-DM recomendamos a consulta do Anexo A do presente trabalho.

3.3.2. SEMMA

A metodologia *SEMMA* foi desenvolvida pelo Instituto *SAS* (SAS Institute Inc.), organização cuja missão é o desenvolvimento de soluções para as áreas de Estatística, Análise de Dados, *Business Intelligence*, *Data Mining* e Suporte à Decisão (SAS, 2009).

Esta metodologia integra-se perfeitamente com o software *Enterprise Miner* (EM) desenvolvido também pelo Instituto *SAS*. Esta aplicação fornece vários tipos de algoritmo de *Data Mining* entre as quais: regressão, classificação, *clustering*. Acrescentado também o facto de permitir também a análise estatística. A característica que sobressai em relação à concorrência é a sua variedade de ferramentas de análise estatística, que são baseadas na longa história e experiência do *SAS* no mercado de análise estatística.

O *Enterprise Miner* contém um conjunto de ferramentas de análise que podem ser combinadas de modo a criar e comparar múltiplos modelos (ver figura 3.5.).

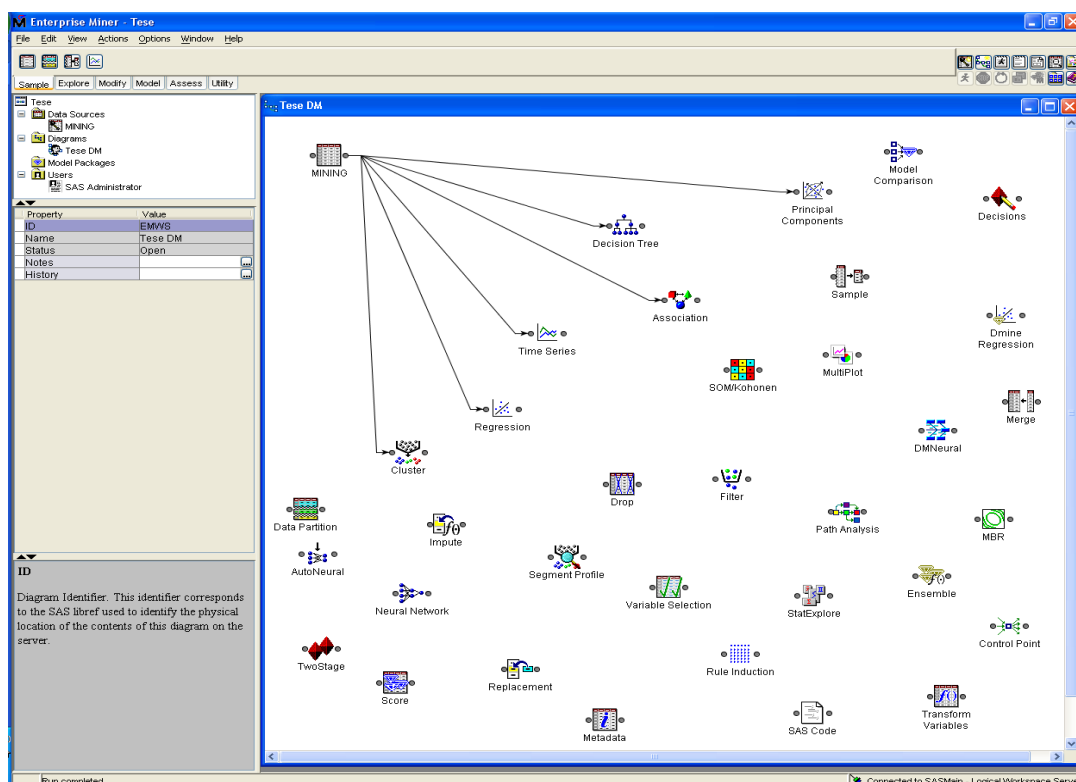


Figura 3.5. – Software SAS Enterprise Miner 5.2

Esta ferramenta possui uma excelente interface gráfica que permitem uma análise rápida dos resultados obtidos.

O Instituto *SAS* define o DM como um processo de extrair informação valiosa e relações complexas de grandes volumes de dados. Com base neste conceito genérico, o processo de *Data Mining* foi dividido em 5 etapas, que compõe o acrónimo *SEMMA*: *Sample* – Amostragem; *Explore* – Exploração; *Modify* – Modificação; *Model* – Modelação; e *Assessment* – Avaliação (Groth, 2000), sendo que cada etapa tem funções específicas e distintas.

O principal objectivo destas cinco etapas (ou estágios) é a obtenção de um modelo analítico que nos permite responder à questão de negócio em análise (Alexandre e Sousa, 2007).

No que diz respeito a etapa Amostragem (*Sample*) temos a possibilidade de realizar uma amostra, extraindo uma parte dos dados que se considera possuir informação significativa e de realizar uma partição dos dados. Na etapa Exploração (*Explore*) irá existir uma exploração dos dados, uma pesquisa de relações antecipadas, de tendências e de anomalias de forma a obter uma maior compreensão e ideias (Alexandre e Sousa, 2007). Após a exploração dos dados tornar-se necessário proceder à modificação dos dados (*Modify*) através da criação, selecção e transformação de variáveis, com o objectivo de focalizar-se no processo de selecção de modelo. A Modelação (*Model*) dos dados é concretizada pela aplicação das ferramentas de análise disponíveis no pacote, nomeadamente árvores de decisão, redes neuronais, entre outras. É esta a etapa responsável pela descoberta de informação, propriamente dita. Por último, a etapa da avaliação dos dados (*Assessment*), gerado de forma a aferir o seu desempenho. Normalmente corresponde à apresentação de um conjunto de dados (*i.e.*, conjunto de teste) ao modelo, avaliando a resposta do mesmo perante casos que não foram utilizados na fase de treino.

Estas etapas, distintas, correspondem a um ciclo, em que as tarefas internas podem ser executadas de forma repetida sempre que se verifique necessário (como ilustrado na figura 3.6.).

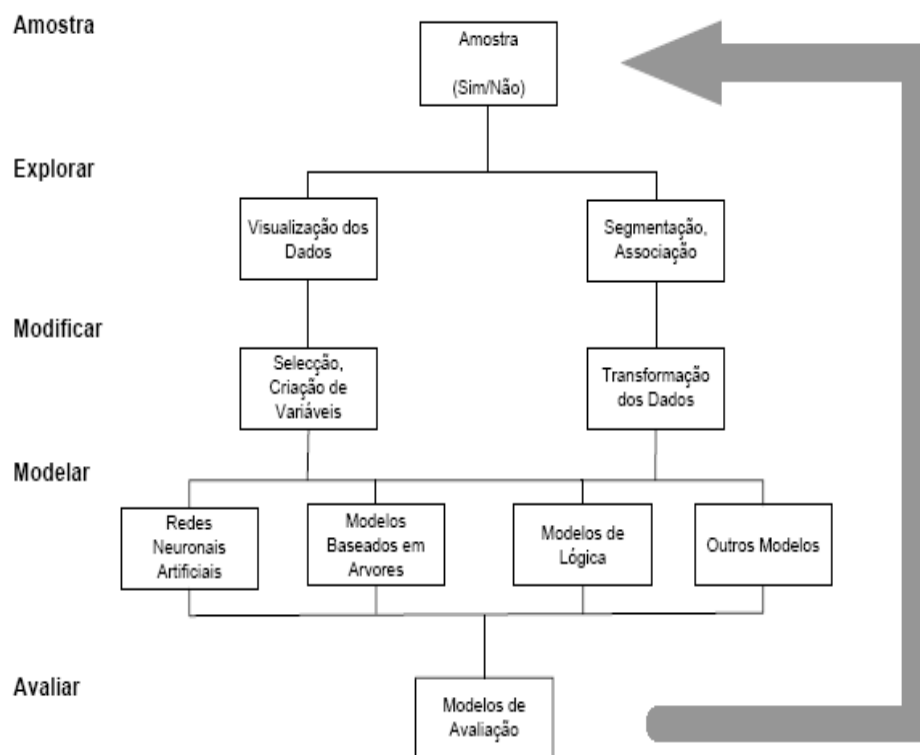


Figura 3.6. – Etapas da metodologia SEMMA

Esta metodologia permite: *i)* uma maior facilidade na aplicação de técnicas estatísticas e de visualização; *ii)* seleccionar e transformar as variáveis explicativas/independentes mais significantes; *iii)* modelar as variáveis para prever os resultados; e *iv)* confirmar a exactidão de um modelo (Alexandre e Sousa, 2007)

Mais do que uma metodologia de *Data Mining*, é considerada um auxiliar na condução de um projecto em todas as suas etapas, desde a especificação do problema até à sua implementação, disponibilizando uma estrutura para a concepção, criação e evolução dos projectos de *Data Mining*, de forma a apresentar soluções para os problemas do negócio.

Para mais informações, encontra-se no Anexo B um resumo mais detalhado do processo SEMMA.

3.3.3. CRISP-DM VS SEMMA

O processo SEMMA é similar à metodologia CRISP-DM compartilhando a mesma essência, estruturado em fases que se encontram inter-relacionado entre si, convertendo o processo de *Data Mining* num processo iterativo e interativo.

Como já foi referido anteriormente, a construção do SEMMA começa com a amostragem dos dados da população para criar um conjunto de dados maneável para a análise, e em seguida explora os dados visualmente para determinar que tipos de padrões e de tendências podem ser encontrados. A seguir, os dados são manipulados para segurar a integralidade e a qualidade dos dados. No módulo da Modelação são usadas técnicas de modelagem apropriadas de estatística tais como a análise de regressão, redes neurais, métodos de raciocínio baseado em árvores de decisão e métodos de séries temporais para descobrir as origens/raízes causais dos padrões e tendências. O passo final é a validação onde é utilizado os modelos desenvolvidos (com os dados iniciais de treino) para ser comparados com a amostra “*holdout*”²³ no fim de determinar a eficácia dos modelos de previsão.

Em relação ao modelo CRISP-DM começa com a compreensão do problema que inclui um foco nos objectivos / metas do negócio e um plano de projecto. Tanto a recolha de dados como a exploração (ou prospecção) foram combinadas num módulo denominado como *compreensão dos dados*. A etapa seguinte trata-se da *preparação dos dados* no qual incorpora actividades como a formatação dos dados, a limpeza de dados e a integração de dados. Os dois últimos módulos focalizam-se sobre a avaliação do resultado do modelo e da implantação dos modelos “em produção”.

Ao analisar estas duas metodologias, podemos verificar que muitas das etapas no processo SEMMA estão directamente correlacionadas com as etapas na metodologia CRISP-DM (ver figura 3.7.).

²³ O método “Holdout”: De acordo com Hill (2002), “a amostra total é dividida aleatoriamente em duas sub-amostras *A* e *B* (o número de casos em cada uma delas deve ser aproximadamente igual). O método “holdout” assume que as duas sub-amostras são representativas da amostra total e do Universo, e portanto, os resultados de uma técnica estatística aplicada à sub-amostra *A* devem ser muito semelhantes aos resultados obtidos por aplicação da mesma técnica à sub-amostra *B*”.

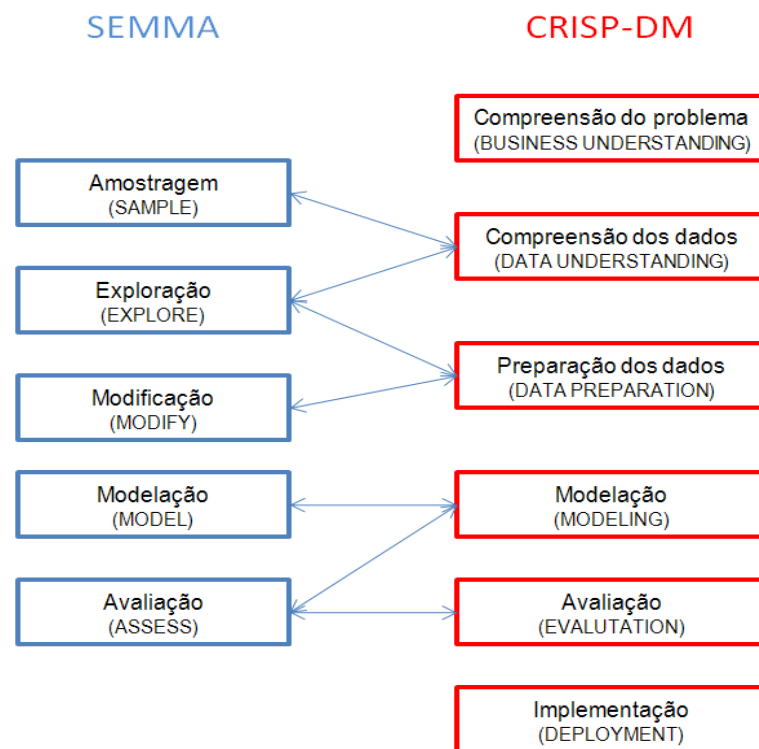


Figura 3.7. – Comparação das interacções entre as fases das metodologias SEMMA e CRISP-DM

No entanto, existem diferenças entre a metodologia SEMMA e a metodologia CRISP-DM que radicam na relação com as ferramentas comerciais. a metodologia SEMMA é perfeitamente integrado com o *SAS Enterprise Miner* (Giudici, 2003). O *SAS Enterprise Miner* oferece uma interface gráfica com a qual podem ser construídos diagramas de projectos de DM, tornando-se numa das grande vantagem em utilizar o SEMMA, pelo facto de se encaixar tão habilmente com as ferramentas da *SAS*.

Enquanto ao CRISP-DM, a sua grande vantagem é de ser livre e gratuito, e ainda pelo facto de ser particularmente detalhado e útil. Salientamos também que o CRISP-DM foi recentemente objecto de uma tentativa de revisão que não produziu aparentemente resultados práticos²⁴.

²⁴ www.crisp-dm.org.

CAPÍTULO IV

*“The roots of education are bitter,
but the fruit is sweet.”*

(Aristóteles)

Clustering

Uma visão natural da segmentação é “tentar determinar quais são as observações de um conjunto de dados que naturalmente devem estar juntos” (Forsyth e Ponce, 2003). Ou seja, trata-se de um problema conhecido como *clustering*.

O termo *clustering* (ou agrupamento) pode ser definido como um processo de agrupamento de objectos (indivíduos, produtos, etc.) em subgrupos (clusters) com características semelhantes (Basu *et al.*, 2009)

Cada grupo, denominado *cluster*, é determinado de forma a obter-se um elevado grau de homogeneidade dentro dos clusters e um alto nível de heterogeneidade entre eles.

Concretamente, para efectuar a segmentação de clientes, optámos por adoptar uma abordagem de *clustering* pelo facto de haver uma grande quantidade de dados, que, tratados com métodos automáticos, poderão permitir a identificação de perfis de clientes anteriormente desconhecidos.

De seguida, discutir-se-á em mais pormenor uma das técnicas de *Data Mining* que é o *clustering*. Primeiro especifica-se o problema de *clustering* (Secção 4.1). Em seguida, descreve-se os objectivos (Secção 4.2), as suas aplicações com ênfase na segmentação (Secção 4.3), sendo ainda abordado a relação entre o *clustering* e a segmentação (Secção 4.4). A seguir, vamos descrever os métodos de *clustering* (Secção 4.5). Finalmente, vamos abordar o tema da avaliação de *clustering* (Secção 4.6).

4.1. Definição do Problema de Clustering

Nesta secção vamos inicialmente definir de modo formal e informal o conceito de *clustering*, em seguida vamos descrever as dificuldades dos problemas de *clustering*, que motivam a existência de vários algoritmos e a necessidade de seleccionar o mais adequado para o problema.

Vamos em primeiro lugar definir e formalizar alguns dos vários termos matemáticos usados no *clustering*, no fim de familiarizar e entender melhor os termos usados neste capítulo (ver tabela 4.1.).

Tabela 4.1. – Descrição de termos matemáticos usados no *Clustering*

Termos	Descrição
Algoritmo	Conjunto de regras ou procedimentos para realizar uma tarefa (Hair <i>et al.</i> , 1998)
Cluster	É um subconjunto de objectos, observações ou exemplos com características afins entre si e distintas dos restantes exemplos fora desse cluster.
Clustering	O clustering é usado com 2 significados: 1) (verbo) processo de geração de clusters a partir dos dados e 2) (nome) resultado do processo de <i>clustering</i> . No plural significa “o resultado de vários processos de geração de clusters”.
Objecto	Um objecto é definido por um conjunto de variáveis

Existem na literatura várias terminologias com significados semelhantes para *clustering*, tais como aprendizagem não supervisionada (Jain e Dubes, 1988), aprendizagem pela observação (Michalski e Stepp, 1983), quantização vectorial (Oehler e Gray 1995) e análise de clusters²⁵ (proposto por Tryon, 1939).

Concretamente, o *clustering* refere-se a agrupar objectos em grupos (clusters), de tal modo que seja minimizada a distância/variação intra-grupo e seja maximizada a distância/variação inter-grupos (Hruschka e Ebecken, 2003).

O *clustering* é um problema difícil (Jain *et al.*, 1999; Kantardzic, 2003; Law *et al.*, 2004) devido, por exemplo: *i*) à diversidade da forma dos dados; *ii*) das suas dimensões; *iii*) da definição do número de clusters (Law *et al.*, 2004); *iv*) da contaminação dos dados com ruído que pode criar distorções nos clusters; *v*) da sua subjectividade, isso porque o mesmo conjunto de dados precisam frequentemente de ser

²⁵ O termo “análise de cluster” é o mais comum e o mais referenciado na literatura.

dividido de forma diferente para diferentes fins (Jain *et al.*, 1999), e *vi*) da escolha do algoritmo.

A escolha do algoritmo de *clustering* é importante, pelo facto de ser empregue, por exemplo, na descoberta de grupos de clientes com valores similares (Böttcher *et al.* 2009). No entanto, a escolha de um algoritmo de *clustering* é difícil pelo facto de existirem na literatura centenas de algoritmos de *clustering*. Sendo que até hoje “não foi possível identificar um algoritmo capaz de tratar adequadamente todos os tipos de dados” (Harrison, 1998).

Fasulo (1999) afirma mesmo que “a escolha de um algoritmo de *clustering* para um determinado problema pode ser uma tarefa intimidante”, devido ao facto do *clustering* ter sido utilizado em várias disciplinas em que cada disciplina utilizou diferentes terminologias e definições básicas para descrever as suas abordagens, dificultando a procura da abordagem mais relevante para um determinado problema.

4.2. Objectivos do Clustering

Os objectivos gerais do *clustering* são:

- 1) Identificar grupos distintos (clusters) de um conjunto de dados em grupos homogéneos, e que sejam diferentes entre si (Kogan *et al.*, 2006);
- 2) Dividir os dados em subclasses "naturais" (Law *et al.*, 2004; Collica, 2007), isto é, que sejam, úteis e apropriados sob o ponto de vista da aplicação;
- 3) Detectar dados não comuns (outliers).

Uma das características que torna o *clustering* uma das técnicas mais utilizadas é a sua capacidade de identificar estruturas directamente dos dados, sem que haja um conhecimento prévio das mesmas. Mas a grande vantagem do uso das técnicas de *clustering* é que ao agrupar dados, podemos *i*) descrever de forma mais eficiente e eficaz as características dos diversos grupos (Collica, 2007); *ii*) ter um maior entendimento da população que os dados representam; e *iii*) possibilitar o desenvolvimento de esquemas de classificação para novos dados (Kantardzic, 2003).

4.3. Aplicações do Clustering

O *clustering* foi já utilizado em várias áreas, tais como: *Biologia* - divisões celulares do ovo para formar o embrião (e.g. de Queiroz e Good, 1997), *CRM - Customer Relationship Management* (e.g. Roussinov e Zhao, 2004), *Psicologia* (e.g. Robinson, 1966; Kremers *et al.*, 2004), *Processamento de imagem* (e.g. Matsuyama, T. 1989; Clément e Thonnat, 1993), *Medicina* (e.g. Biehl *et al.*, 2007), *Processamento de sinal* (e.g. Shekhar *et al.*, 1994) e *Marketing (research)* (e.g. Mahajan e Jain, 1978; DeSarbo, 1982; Punj e Stewart, 1983; DeSarbo e Mahajan, 1984; Arabie e Hubert, 1994).

No marketing aplicou-se a análise de clusters na segmentação de mercados (Punj e Stewart, 1983) e também para determinar os mercados-alvo, para posicionar o produto e desenvolver novos produtos (Kotler *et al.*, 2005). Mais recentemente foi aplicada à selecção de testes de mercados (ver técnicas experimentais). No entanto, no *marketing (research)*, a análise de clusters é usada não só para exploração dos dados, sendo geralmente usada para a identificação e definição de segmentos de mercado tornando-se o centro da estratégia de *marketing* da empresa (Wedel e Kamakura, 1999).

Historicamente o uso da análise de cluster foi frequentemente visto com cepticismo (Punj e Stewart, 1983). Por exemplo vários autores (e.g. Shuchman, 1967 *apud* Punj e Stewart, 1983; Neidell, 1970) expressaram alguma preocupação quanto ao uso dessa ferramenta. Mas hoje em dia, “todos” concordam que a análise de clusters fornece um grande leque de oportunidades como, por exemplo, no *marketing* (Punj e Stewart, 1983).

Han e Kamber (2006) destacam que, na área empresarial, o clustering pode ajudar a descobrir grupos distintos nas bases de clientes e caracterizar os grupos de clientes baseado nos padrões de compras. O interesse para a análise de cluster tornou-o naturalmente uma ferramenta popular para a segmentação de mercado (Athanassopoulos, 2000; Van Raaij *et al.*, 2003; Mentzer *et al.*, 2004).

De acordo com Wedel e Kamakura (1999), o número de aplicações de métodos de agrupamentos (*clustering*) no *marketing* é “enorme tornando virtualmente impossível a revisão de todas as aplicações até a data”. Para além disso, podemos acrescentar o facto que em várias investigações sobre a segmentação, os autores não especificaram

que tipo de método de *clustering* foi aplicado e sabendo que “existe diferentes abordagens de análise de cluster para a segmentação de mercado” (Zakrzewska e Murlewski, 2005), torna a tarefa da revisão da literatura ainda mais difícil.

4.4. Clustering & Segmentação

Em termos de terminologias, a segmentação refere-se, frequentemente ao processo, metodologia ou estratégia, enquanto o *clustering* refere-se, em geral, ao método (algoritmo) utilizado para o agrupamento. No entanto, é muito comum encontrar os termos *clustering* e segmentação associados na literatura (Tan *et al.*, 2006).

Na teoria, existe uma diferença entre *clustering* e segmentação. O clustering é o agrupamento de observações usando métodos automáticos e é muito genérico. Enquanto a segmentação é uma técnica usada em *marketing* para efectuar o agrupamento de clientes (ou divisão de mercados em subgrupos), e pode ser manual e / ou (semi-) automático. Os autores Tan *et al.* (2006) salientam ainda que a segmentação distingue-se também do *clustering* “ao usar frequentemente técnicas simples” (e.g. dividir a população em relação ao seu rendimento). No entanto, na “prática”, no que diz ao processo de selecção, esses dois termos são muito semelhantes (ver figura 4.1.).

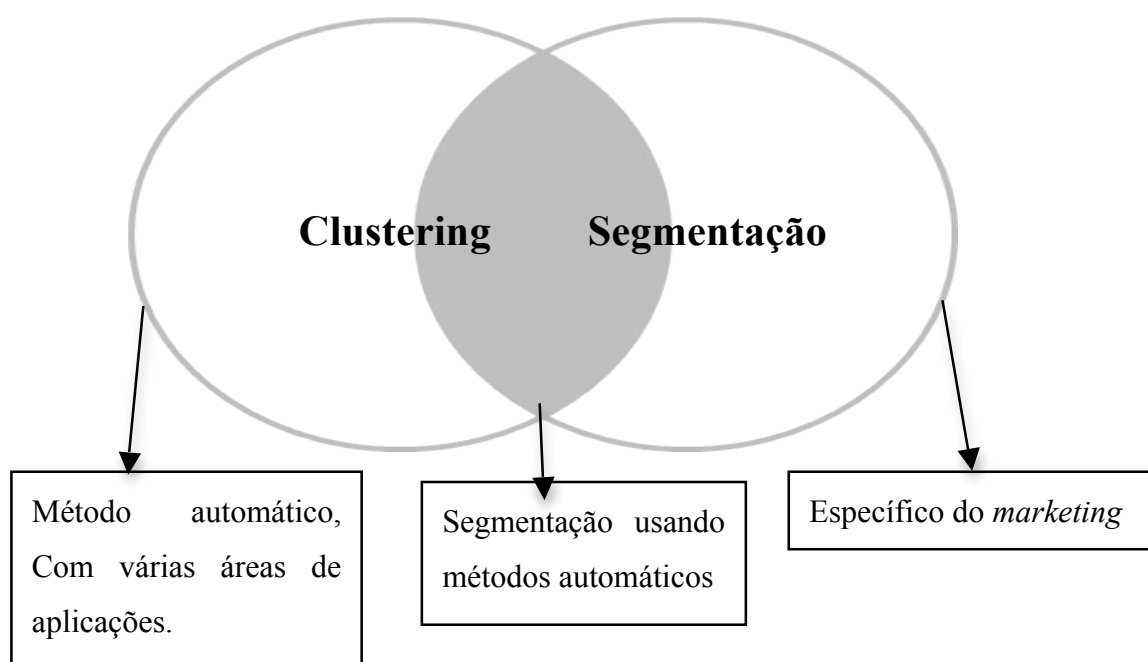


Figura 4.1. – Clustering & Segmentação

No caso concreto da segmentação em *marketing*, o termo mais comum para descrever o *clustering* é “análise de clusters”, sendo mesmo o mais referenciado na literatura (Wedel e Kamakura, 1999). Mas no caso de se referir aos algoritmos utilizados para o agrupamento é mais frequente o uso do termo *clustering*. Por exemplo, os autores Aldenderfer e Blashfield (1984) referem-se ao *clustering* como “o algoritmo tipicamente utilizado para o guiar a partir dos dados na segmentação de mercado”.

De acordo com Hand *et al.* (2007) é possível distinguir esses termos com base na diferença de objectivos. Na segmentação, o objectivo é simplesmente de “dividir os dados de uma forma conveniente” (Collica, 2007), enquanto que no *clustering*, o objectivo é, como já foi referido anteriormente, de dividir os dados em subclasses “naturais” ou grupos (*idem*).

Para entender melhor o que faz o *clustering*, podemos ilustrar como exemplo (ver figura 4.2.) um caso de segmentação realizado com métodos automáticos que identificou nos dados características comuns (*e.g.* indivíduos, produtos, etc...) e agrupou essas características em quatro clusters.

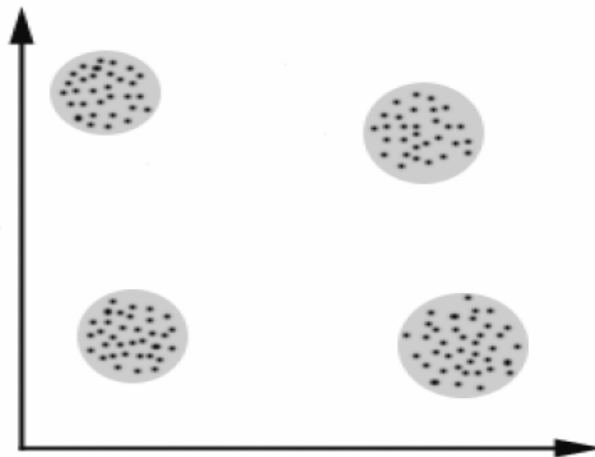


Figura 4.2. – Exemplo de Clustering / Segmentação

No entanto, é possível realizar vários agrupamentos a partir do mesmo conjunto de dados. A figura 4.3. ilustra, por exemplo, três diferentes maneiras de dividir o mesmo conjunto de dados (ver figura 4.3a) em vários grupos homogêneos (clusters).

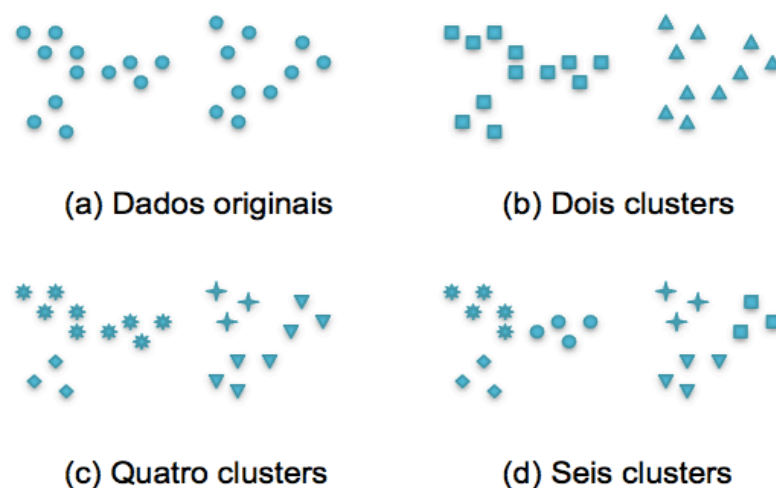


Figura 4.3. – Exemplo de diferentes possibilidade de *clustering* a partir do mesmo conjunto de dados (Adaptado de Tan *et al.*, 2006)

Convém referir que a definição do agrupamento desejado (*clustering*) depende da natureza dos dados e dos resultados desejados (Tan *et al.*, 2006). Para realizar os agrupamentos são usados métodos (algoritmos) de *clustering*, apresentados a seguir.

4.5. Métodos de Clustering

Nesta secção, vamos discutir duas dimensões que são em primeiro lugar os tipos de agrupamentos e em segundo lugar os diferentes tipos de algoritmos existentes na literatura.

Relativamente aos métodos de *clustering*, não há uma classificação simples dos algoritmos de *clustering* (Jain e Dubes, 1988; Kaufman e Rousseeuw, 1990; Duda *et al.*, 2002). No entanto, uma proposta interessante para classificar os diferentes métodos baseando-se nos padrões de sobreposição dos grupos obtidos foi sugerida por Hruschka (1986) no qual distingue três grandes formas de agrupamento: *Sobrepostos (Overlapping)*, *Não Sobrepostos (Nonoverlapping)* e *Difusos (Fuzzy)* (ver figura 4.4.), tendo em conta a configuração da matriz de **P**ertença aos segmentos que produzem (Wedel e Kamakura, 1999).

Relativamente a função da matriz, a matriz **P** possui N linhas, correspondendo aos consumidores, e S colunas, correspondendo aos segmentos produzidos (Wedel e

Kamakura, 1999). As entradas na matriz indicam a afectação dos consumidores aos segmentos.

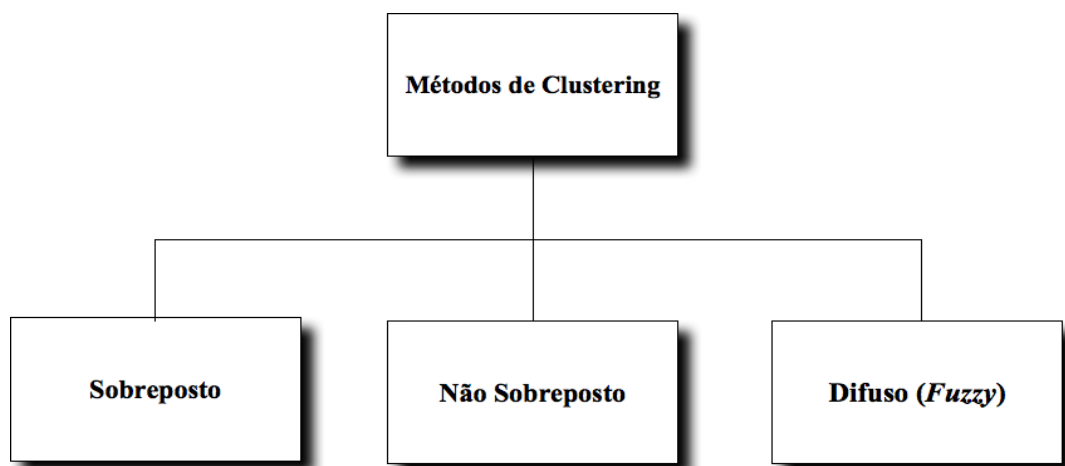


Figura 4.4. – Classificação dos algoritmos de clustering (Adaptado de Hruschka (1986), Wedel e Kamakura, 1999; Oliveira-Brochado e Martins, 2008)

De acordo com Wedel e Kamakura (1999) o agrupamento não sobreposto assume que um objecto (consumidor) pertence a um e a um só segmento; deste modo, a matriz $\mathbf{P}$ possui em cada linha um elemento igual a um, sendo os restantes elementos iguais a zero. Neste contexto destacam-se os tradicionais métodos de agrupamento (*clustering*) que podem ser dividido em dois tipos básicos: o método de *clustering* hierárquico (Sokal e Sneath, 1963; King 1967;) e o método partitivo ou não hierárquico (Everitt, 1993).

Os métodos pertencentes às classes de agrupamento sobreposto e difuso (*fuzzy*) relaxam o pressuposto de que a fronteira de cada grupo é bem definida. No caso dos grupos sobrepostos, um consumidor pode pertencer a mais do que um segmento, *i.e.*, uma linha na matriz $\mathbf{P}$ pode possuir várias entradas iguais a um (Wedel e Kamakura, 1999). No caso dos métodos difusos, um elemento pode pertencer em graus diferentes a diferentes clusters.

A abordagem difusa (*fuzzy*) fornece valores de pertença que se encontram entre 0 e 1. Estas pertenças permitem a identificação de *pontes*, *i.e.*, elementos que se localizam em duas das classes (Oliveira-Brochado e Martins, 2008).

De um modo geral, um bom método de *clustering* deve produzir clusters de alta qualidade, assegurando uma alta similaridade intra-clusters e uma baixa similaridade inter-cluster. Neste contexto, vários autores para avaliar os clusters não hierárquicos

baseiam-se numa formula tradicional que é a seguinte:

$$\text{Qualidade do Cluster} = \frac{\text{Distância Inter - Cluster}}{\text{Distância Intra - Clusters}}$$

No que diz respeito à avaliação da qualidade de cluster hierárquicos é necessário ter também em conta a dimensão dos clusters.

4.5.1. Distância entre Exemplos

Um conceito fundamental na utilização das técnicas (métodos) de agrupamento (*clustering*) é a escolha de um critério que mede a distância entre dois exemplos (ou observações ou objectos ou casos ou indivíduos) (Jain *et al.*, 1999).

Han e Kamber (2006) destacam que a mais utilizada é a distância *Euclidiana*, que é a distância em linha recta entre os dois pontos que representam os objectos. Considerando objectos com apenas dois atributos i e j , temos na figura 4.5. a representação da distância *Euclidiana*.

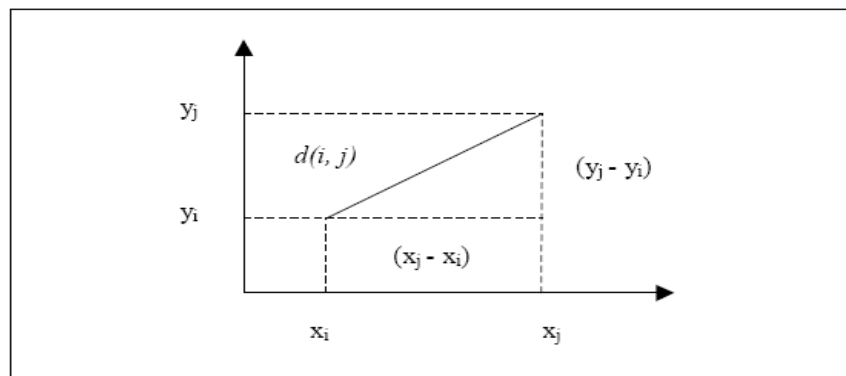


Figura 4.5. – Exemplificação gráfica da Distância *Euclidiana*

A distância *Euclidiana* entre dois atributos é expressa na seguinte fórmula:

$$d(i, j) = \sqrt{(y_j - y_i)^2 + (x_j - x_i)^2}$$

Por exemplo, no caso de uma matriz de dissimilaridade cada elemento da matriz representa a distância entre pares de objectos. Aqui, considerando n objectos a serem agrupados, teremos uma matriz quadrada de tamanho $n \times n$ como a que segue:

$$D = \begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & \ddots & \\ d(n,1) & d(n,2) & d(n,3) & \dots & 0 \end{bmatrix}$$

Com $d(i, j)$ representando a distância entre o objecto i e o j .

Tipicamente a similaridade entre dois objectos i e j assume valores entre 0 e 1, onde o valor 0 expressa que os dois objectos não são similares, enquanto que o valor 1 expressa máxima similaridade. Ou seja, as medidas de similaridade ou dissimilaridade são números positivos.

A seguir, vamos discutir os diferentes tipos de algoritmos

4.5.2. Método Hierárquico

Os métodos hierárquicos podem ser aglomerativos (*agglomeration method*) ou divisivos (*divisive method*) (Jain e Dubes, 1988; Kaufman e Rousseeuw, 1990). No primeiro, cada objecto ou observação parte como sendo um cluster e nos passos subsequentes os dois objectos mais próximos vão-se agregando num só cluster. O divisivo é um processo inverso, em que se parte de um só grupo que inclui todos os indivíduos e através de divisões sucessivas e sistemáticas, as observações mais afastadas vão sendo retiradas e constituem-se clusters mais pequenos (Hair *et al.*, 1998; Reis, 2001). A aplicação desse método permite apresentar os resultados sob a forma de dendograma (Jain e Dubes, 1988) e/ou árvores de decisão (Johnson, 1967).

As cinco estratégias (ou algoritmos) mais populares do método aglomerativo são: *i)* o critério do Vizinho mais próximo - *Single-Linkage Method* (Sokal e Sneath, 1963) ; *ii)* o critério do Vizinho mais afastado - *Complete-Linkage Method* (King,

1967); iii) o critério da Média dos grupos (*Average-Linkage Method*); o critério do Centróide - *Centroid Method*; e iv) o critério de Ward - *Ward's Method* (Ward, 1963).

O *Single Linkage* é um dos algoritmos de *clustering* hierárquico mais simples descrito inicialmente por Sokal e Sneath (1963) e Johnson (1967). Esse método utiliza a técnica do vizinho mais próximo, na qual a distância entre dois clusters é determinada pela distância do par de exemplos mais próximos. O *Complete Linkage* utiliza uma técnica conhecida como o vizinho mais distante. Ao contrário do algoritmo *Single Linkage*, esse algoritmo determina a distância entre dois clusters de acordo com a maior distância entre um par de exemplos, sendo que cada exemplo pertence a um cluster distinto.

Em relação a distância entre clusters, a medida é realizada computacionalmente através de um função de distância.

Esta função recebe dois elementos e devolve o valor correspondente à “semelhança” entre os dois elementos (Han e Kamber, 2006). No entanto, cada aplicação tem uma noção diferente de semelhança, pelo que não existe uma função de distância adequada a todos os domínios.

Ou seja, o *Single Linkage* e *Complete Linkage* representam dois extremos em termos de distância entre cluster. Como todos os procedimentos que envolvem esses extremos, tendem a ser altamente sensíveis à presença de outliers. Assim, o uso de uma abordagem intermédia é um caminho natural para amenizar esse problema (Duda *et al.*, 2002). No algoritmo *Average linkage*, a distância entre dois clusters é definida como a média das distâncias entre todos os pares de exemplos em cada cluster, sendo que cada par é composto por um exemplo de cada cluster (ver figura 4.6.).

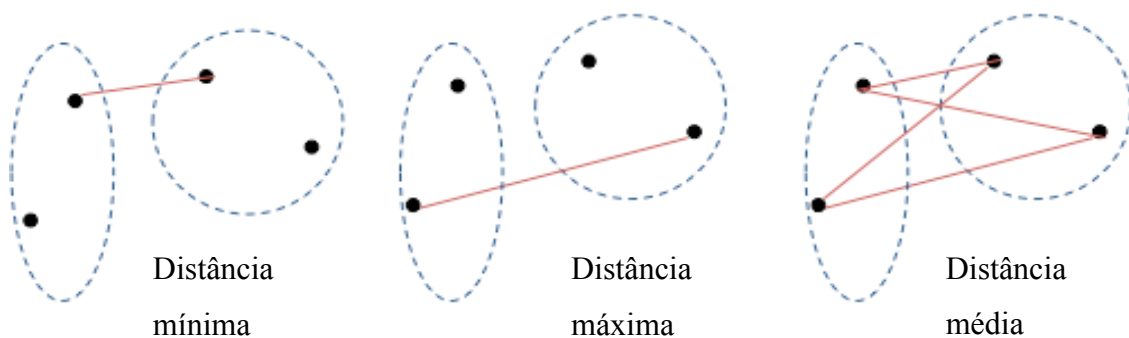


Figura 4.6. – Métodos *Linkage*

O algoritmo *Centróide* calcula o ponto central (*centróide*²⁶) do conjunto em causa, depois para avaliar a pertença de um determinado elemento a um conjunto é avaliada a distância entre o elemento e os centróides existentes (ver figura 4.7.). Com algoritmo centróide, os grupos identificados são representados por um vector que armazena a média de cada atributo. Assim, a distância entre os clusters é definida em termos de distância no espaço euclidiano entre o representante de cada cluster. Uma desvantagem desse algoritmo está relacionada ao tamanho dos clusters seleccionados para a junção, pois quando são muito diferentes, o centróide do novo cluster será semelhante ao centróide do cluster de maior tamanho, o que acarretará na perda das características do menor cluster (Duda *et al.*, 2002).

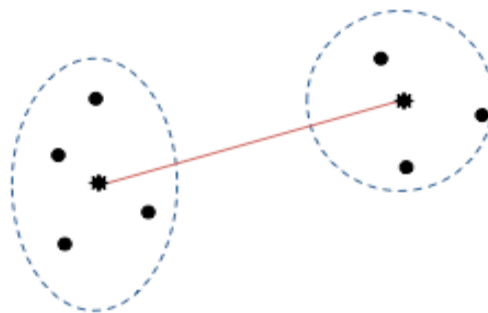


Figura 4.7. – Método do centróide

Os três algoritmos *Linkage* descritos anteriormente operam sobre a matriz de similaridade e não necessitam acesso directo aos valores dos atributos dos exemplos, enquanto o algoritmo baseado no centróide deve ter acesso directo aos exemplos.

O método de *Ward* (Ward, 1963) é um método hierárquico aglomerativo, uma vez que, à partida, há tantos clusters como objectos. Este algoritmo foi elaborado de modo a que as partições formadas minimizem a perda associada a cada agrupamento. Estes grupos são fundidos, sucessivamente, com base nas suas semelhanças, até se formar apenas um cluster. Nesse procedimento, todas as possíveis uniões entre pares de clusters são consideradas e os clusters que apresentam a mínima perda de informação são seleccionadas para o agrupamento.

²⁶ O centróide de um *cluster* é o ponto cujos valores dos parâmetros são a média dos valores dos parâmetros de todos os pontos do *cluster*.

Em geral, o método *Ward* é considerado como um método eficiente, mas o inconveniente que mais frequentemente lhe é apontado é a formação de *clusters* de dimensão reduzida.

Por concluir, Everitt (1993) salienta que não há um método particular de clustering hierárquico que possa ser preferido em relação aos outros.

4.5.3. Método não Hierárquico

Os métodos não hierárquicos aplicam-se directamente sobre os dados originais e que partem de uma partição inicial dos indivíduos por um número de grupos definidos pelo utilizador.

O k-médias (*k-means*) proposto por MacQueen (1967) constitui um método não hierárquico que aplicado sobre um agrupamento de n objectos procura agrupá-los num número K , previamente determinado, de grupos (Hartigan e Wong 1979). Cada objecto do agrupamento pertence num único agrupamento. Os elementos que compõem os grupos apresentam variância mínima entre si.

Em termos de funcionamento, o algoritmo k-médias passa pelas seguintes etapas:

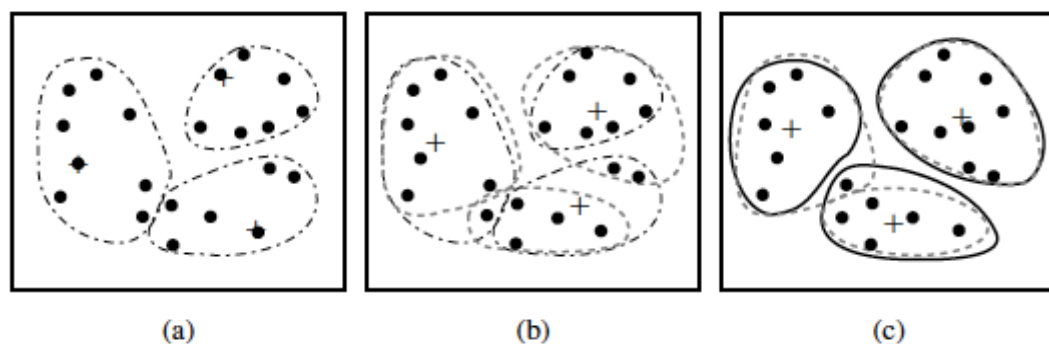
- (1) Escolher aleatoriamente k objectos (pontos) de D que serão os centros dos clusters iniciais – centróides;
- (2) Repetir;
- (3) (re) atribuir cada objecto ao cluster do qual o objecto está mais próximo, em termos de similaridade ao valor médio dos objectos no cluster;
- (4) actualizar a média do cluster, isto é, calcular o valor médio dos objectos para cada cluster;
- (5) até não haver alterações.

Onde k representa o número de clusters (parâmetro); e D representa um conjunto de dados que contém n objectos (Han e Kamber, 2006).

Para exemplificar melhor o processo do método k-médias, baseamo-nos num exemplo apresentado pelos autores Han e Kamber (2006). Neste exemplo foram

escolhidos aleatoriamente três objectos como centros iniciais dos clusters, representados pelo símbolo “+”. Cada objecto é atribuído a um cluster, com base no centro do cluster ao qual é o mais próximo, apresentado na figura 4.8(a). A seguir, os centros dos cluster são actualizados. Usando os novos centros dos clusters, os objectos são redistribuídos nos clusters em que o centro do cluster é o mais próximo, como ilustrado na figura 4.8(b). Este processo repete-se, conduzindo a figura 4.8(c).

Figura 4.8. – Método k-médias (A média de cada cluster é representado pelo símbolo “+”).



Fonte: Han e Kamber (2006)

No entanto, existem várias limitações do método k-médias (Han e Kamber, 2006) destacando as seguintes: Em primeiro lugar, o método é aplicável apenas a dados numéricos, por ser necessário calcular o valor médio de cada variável. Em segundo lugar, o método k-médias é adequado principalmente para a descoberta de clusters com forma convexa e com dimensões semelhantes. Finalmente, o método k-médias é muito sensível ao ruído e aos *outliers*. Tais dados podem influenciar substancialmente o valor médio, e produzir resultados errados.

4.5.4. Métodos Fuzzy (difuso)

O conceito "fuzzy" (Zadeh, 1965) pode ser entendido como uma situação que admite valores lógicos intermediários entre o “Sim” (Verdadeiro) ou “Não” (Falso) como o “talvez”. Neste método é criada uma função que indica o grau de pertença de cada elemento, variando entre 0 e 1. Ou seja, o conceito *fuzzy* e o grau de pertença

foram introduzidas para evitar a passagem brusca de uma classe para outra e autorizar os elementos a não pertencer completamente nem o uma nem ao outro ou ainda pertencer parcialmente a cada uma (Tam *et al.*, 2003).

No método difuso (fuzzy) é necessário definir o grau de homogeneidade que se deseja atingir na sequência do processo de agrupamento. Um grau de homogeneidade máximo conduzir-nos-á a considerar cada indivíduo como um cluster, solução muito pouco operacional e difícil de pôr em prática. Um grau mínimo de homogeneidade produziria um único cluster onde reencontrar-se-iam todos os indivíduos, solução igualmente insatisfatória (Tam *et al.*, 2003). A medida da homogeneidade pode aumentar o grau de precisão da segmentação, tendo como consequência a redução da dimensão do mercado (os indivíduos que devem satisfazer às exigências mais fortes para fazer parte de um cluster). Ou seja, o decisor / analista pode exigir uma selecção mais precisa do seu mercado (um grau mais forte de homogeneidade no cluster), o que implica provavelmente um crescimento da taxa de resposta positiva por parte dos indivíduos, mas também um mercado de menor dimensão. Encontrar um compromisso satisfatório entre “o grau de homogeneidade e a dimensão do mercado-alvo representa portanto uma pergunta fundamental na elaboração da estratégia marketing” (Tam *et al.*, 2003).

Podemos citar como exemplo, o método de clustering *Fuzzy c-means* (FCM), desenvolvido por Dunn (1974) e revisto por Bezdek (1981), que é bastante popular.

Recentemente foram desenvolvidos um grande número de novos métodos de clustering, alguns dos quais são mencionados em Xu e Wunsch (2005).

4.6. Avaliação de Métodos de Clustering

A maior parte dos métodos de avaliação de clustering são baseados em medidas de semelhança ou de diferenças entre objectos (dissimilaridade) (Jain e Dubes, 1988). A medida de semelhança (ou proximidade) pode ser definida como a medida de similaridade (medida de igualdade entre dois objectos) ou dissimilaridade (medida de diferença entre dois objectos) entre os dados (Kononenko e Kukar, 2007).

A partir da definição do critério de semelhança é construída uma matriz de

similaridade²⁷ ou de dissimilaridade (Kononenko e Kukar, 2007) como ponto de partida do método de agrupamento.

Rand (1971) afirma que, para avaliar o resultado dos métodos de *clustering*, é importante ter em conta as seguintes questões:

- i) Qual a capacidade do método de identificar os clusters “naturais”?
- ii) Qual é a sensibilidade de um método face à perturbação nos dados?
- iii) Qual é a sensibilidade do método a pequenas diferenças nos conjuntos de objectos?; e
- iv) Na utilização de dois métodos, os resultados produzidos são diferentes quando aplicado aos mesmos dados?.

No entanto, as diferentes técnicas de *clustering* podem produzir soluções diferentes, e a mesma técnica pode mesmo produzir um resultado diferente com o mesmo conjunto de dados (Punj e Stewart, 1983; Jain e Dubes, 1988).

Segundo Han e Kamber (2006) existem características exigidas para avaliar os métodos de *clustering* que são as seguintes: i) escalabilidade; ii) capacidade de lidar com diferentes tipos de atributos; iii) interpretabilidade e usabilidade dos resultados; iv) capacidade de lidar com um elevado número de dimensões; v) capacidade em lidar com dados com ruídos; vi) descoberta de clusters com formas arbitrárias; vii) sensibilidade à ordem dos registos de entrada; e viii) possibilidade de exprimir restrições (ver Basu *et al.*, 2009).

De acordo com Agrawal *et al.* (1998), nenhuma técnica de *clustering* actual satisfaz todos estes “pontos” adequadamente, embora um trabalho considerável tenha sido feito para atender a cada ponto separadamente.

²⁷ A matriz de similaridade é diferente de uma matriz de correlação. A diferença entre uma matriz de similaridade e uma matriz de correlação é que esta última contém semelhanças entre variáveis (por exemplo: renda e educação), enquanto a matriz de similaridade contém semelhanças entre observações (cartão x, cartão y).

CAPÍTULO V

*“The only source of knowledge
is experience.”*

(Albert Einstein)

Seleccção dos Clusterings

5.1. Definição do Problema

As ferramentas descritas no capítulo IV facilitam a criação de muitos *clusterings* (ou segmentações). Muitos deles são potencialmente interessantes, porque representam perspectivas diferentes sobre um dado mercado. No entanto, outras serão redundantes, porque representam segmentações semelhantes. Escolher segmentações “diferentes” é difícil (Law *et al.*, 2004) e implica um grande investimento de tempo (Hoeck *et al.*, 1998).

O objectivo da selecção de segmentações ou *clusterings*, é o de, partindo de um conjunto de *clusterings* obtidos a partir da mesma base de dados, eliminar os *clusterings* potencialmente redundantes e identificar aquelas que são potencialmente interessantes.

Optámos por desenvolver uma metodologia própria para eliminar *clusterings* redundantes, pois não encontramos nenhuma existente na literatura. Na nossa proposta usamos dois métodos de selecção de *clusterings* (semi-) automáticos: a selecção simples e a selecção usando índices de similaridade. A seguir serão discutidos os respectivos métodos e o processo de selecção de *clusterings* potencialmente interessantes.

Um processo (semi-) automático permite diminuir a quantidade de interacção humana e de reduzir o trabalho repetitivo.

Os pressupostos do trabalho realizado foram:

- (i) há interesse em analisar várias segmentações (*clusterings*) para ter diferentes perspectivas do mercado;
- (ii) são geradas n (potencialmente muito grande) segmentações (*clusterings*) onde pretendemos seleccionar m (pequena) para analisar; e
- (iii) as segmentações/*clusterings* interessantes são as que são diferentes entre si, porque agrupam os clientes de maneiras diferentes, o que permite as tais perspectivas diferentes do mercado.

Neste trabalho, optámos por dividir os “clientes” em poucos clusters uma vez que as segmentações obtidas serão analisadas manualmente. Assim, foram gerados três ($n=3$), cinco ($n=5$) e dez ($n=10$) clusters.

5.3. Metodologia de Selecção de Segmentações (Semi-) Automática Simples

A segmentação (semi-) automática simples, é baseada em matrizes de confusão. Em problemas de classificação, a *matriz de confusão*, também denominada tabela de contingência, compara basicamente, classe por classe, a relação “entre os dados de referência conhecidos e os resultados correspondentes de uma classificação automatizada” (Lillesand *et al.*, 2004). Para a análise dos resultados de dois métodos de *clusterings*, a *matriz de confusão* é uma ferramenta muito útil pois permite colocar dois métodos de *clusterings* frente-a-frente numa só tabela (matriz). Assim, vamos poder detectar os *clusterings* semelhantes e identificar os potencialmente redundantes.

Na tabela 5.1., podemos ver um exemplo de uma matriz de confusão para comparar dois *clusterings* de tamanho 5 e 3. Cada célula $M_{i,j}$ da matriz contém o número de casos que pertencem ao cluster i numa das segmentações e ao cluster j na outra.

Tabela 5.1. – Exemplo de Matriz de confusão

Clustering A (com NC de 5)	Clustering A (com NC de 3)				
	$M_{i,j}$	1	2	3	Σ_i
	1	1160	5	20	1185
	2		414		414
	3	69	4	38	111
	4	426	42	190	658
	5	14	4	2614	2632
	Σ_j	1669	469	2862	

O processo de selecção semi-automático simples tem dois passos. No primeiro compara-se *clusterings* com o mesmo tamanho e no segundo compara-se *clusterings* de tamanhos diferentes.

- (i) Comparação de *clusterings* dois a dois com o mesmo número de clusters – 3x3; 5x5 e 10x10; e
- (ii) Comparação de *clusterings* dois a dois com diferentes números de clusters – 3x5; 3x10 e 5x10.

A divisão do processo de comparação em duas etapas evita comparar todas as segmentações duas a duas. Assim, comparámos primeiro segmentações/*clusterings* com o mesmo número de clusters, onde, em princípio, será mais fácil detectar diferenças e, só depois comparamos segmentações com números diferentes de clusters, tendo em conta os resultados obtidos no primeiro passo.

Na descrição mais detalhada do método vamos seguir um exemplo onde foram utilizados quatro métodos de *clustering* (*A*, *B*, *C* e *D*) com número de clusters (NC) igual a 3, 5 e 10. Assim *D3*, por exemplo, representa o *clustering* obtido pelo método *D* com número de clusters igual a 3. Na primeira etapa analisamos, então, os pares possíveis com o mesmo número de clusters, como se pode ver na tabela abaixo.

Tabela 5.2. – Exemplificação da comparação de *clusterings* por pares com o mesmo número de clusters

3x3
Clustering A3 - Clustering B3
Clustering A3 - Clustering C3
Clustering A3 - Clustering D3
Clustering B3 - Clustering C3
Clustering B3 - Clustering D3
Clustering C3 - Clustering D3

Vejamos agora dois casos simples que ilustram situações distintas. Por exemplo, o *Clustering A3* e o *Clustering B3* são iguais (ver tabela 5.3.). Este exemplo reflecte um caso perfeito de segmentação redundante em que os dois *clusterings* são similares. A diagonal da matriz evidencia a semelhança dos *clusterings*. Um deles pode ser eliminado, simplificando o processo de análise.

Tabela 5.3. – Exemplo de segmentações semelhantes

	Clustering A			
Clustering B	1	2	3	Total Geral
1	4997			4997
2		2		2
3			1	1
Total Geral	4997	2	1	5000

No segundo exemplo, o *Clustering C3* e o *Clustering D3* são bastante diferentes. Neste caso (ver tabela 5.4.), a diferença manifesta-se no facto de a maioria das células estarem preenchidas. Assim sendo, estamos perante *clusterings* significativamente diferentes e potencialmente interessantes.

Tabela 5.4. – Exemplo de segmentações diferentes

	Clustering C			
Clustering D	1	2	3	Total Geral
1	2497	350	453	3300
2	1461	126	112	1699
3			1	1
Total Geral	3958	476	566	5000

Em qualquer um dos passos, para a selecção dos pares tenham ou não o mesmo NC, observamos as duas seguintes medidas:

- (i) Número de células preenchidas (NCP);
- (ii) Valor mais elevado na matriz de confusão (MaxM).

Quanto mais elevado for o valor da primeira medida e mais baixo o da segunda para um determinado par, mais distintos são os *clusterings* desse mesmo par. Depois desta pré-selecção de *clusterings*, a segunda etapa consiste em comparar conjuntos com diferentes números de clusters.

A figura 5.1. ilustra como se obtêm os valores das medidas no exemplo dado.

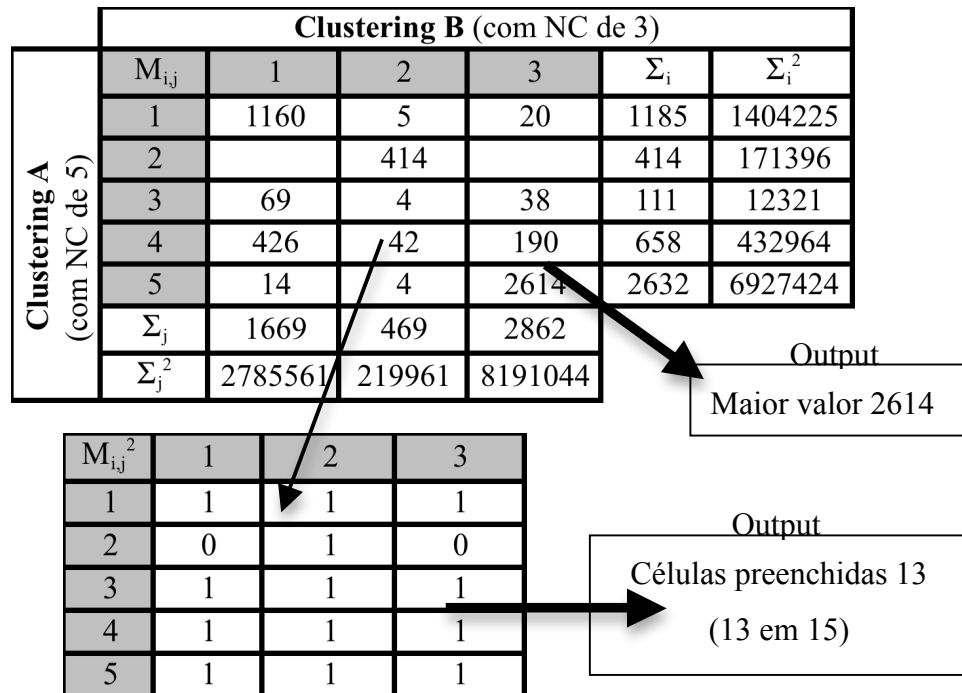


Figura 5.1. – Processo de selecção semi-automática simples com diferentes pares

5.4. Selecção de Segmentações usando Índices de Similaridade

Nesta segunda abordagem para a selecção (semi-) automática de *clusterings* partimos também das matrizes de confusão descritas anteriormente e que são obtidas comparando os diversos *clusterings* dois a dois. A partir de cada matriz de confusão vamos calcular um valor de distância entre os dois *clusterings* respectivos.

Tabela 5.5. – Exemplo de matriz de confusão usado para a comparação dos *clusterings*

Clustering A (com NC de 10)	Clustering B (com NC de 5)						
	M_{ij}	1	2	3	4	5	Σ_i
	1	22	6	6	2	2128	2164
	2	10	1		1	14	26
	3		347				347
	4	91	7	3	27	60	188
	5	281	35	10	74	263	663
	6	314	15	9	48	123	509
	7	429			5	16	450
	8	33	2	3	10	24	72
	9			79			79
	10	5	1	1	491	4	502
	Σ_i	1185	414	111	658	2632	
	Σ_i^2	1404225	171396	12321	432964	6927424	

Vejamos um exemplo. Na tabela 5.5., vemos mais uma vez uma matriz de confusão obtida tal como na primeira abordagem. A coluna e linha com as somas dos quadrados serão justificadas mais abaixo. A ideia chave desta segunda abordagem é a de que dois *clusterings* A e B são semelhantes, se os exemplos que estão no mesmo cluster em A têm grande probabilidade de ficarem no mesmo cluster também em B . Assim, se estiverem em clusters diferentes em A , também estarão em clusters diferentes em B . Assim, a medida de distância que iremos definir passa por verificar que pares de exemplos estão, ou não estão, no mesmo cluster em ambas as segmentações.

Como é então calculada a distância entre dois *clusterings*? Podemos ver como exemplo uma matriz 2×2 , que aqui chamaremos de concordância, apresentada na tabela 5.6., adaptada de Albantineh *et al.* (2006). Nesta matriz comparam-se dois métodos L e C . A linha $L1$ conta os pares que ficam no mesmo cluster em L e a linha $L2$ conta os pares que ficam em diferentes clusters. As colunas $C1$ e $C2$ funcionam de forma análoga para o método C . Ou seja: a é o número de pares que são postos no mesmo cluster de acordo com ambos métodos de *clustering*, b é o número de pares que são postos no mesmo cluster de acordo com o método 1, mas não de acordo com o método 2, c é o número de pares que são postos no mesmo cluster de acordo com o método 2, mas não de acordo com o método 1, d é o número de pares que são postos em clusters diferentes em ambos os métodos (Albantineh *et al.*, 2006).

Tabela 5.6. – Matriz de concordância entre os *clusterings*

Método 1	Método 2		
		C1	C2
		C1	C2
L1	a	b	a + b
L2	c	d	c + d
	a + b	b + d	M

Esta matriz de concordância não é mais do que um resumo da matriz de confusão que compara os dois *clusterings*, e pode ser obtida a partir desta. Jain e Dubes (1998) e Albatineh *et al.* (2006) definiram como calcular os valores a , b , c , e d apresentadas na tabela 5.6, a partir da matriz de confusão M_{ij} .

Assim, para uma base de dados de tamanho m o número total de pares será M , que por sua vez é igual à soma dos valores da matriz de concordância.

$$a + b + c + d = M = \binom{m}{2} = \frac{1}{2}m(m-1)$$

Em que $m = \sum_{i=1}^I \sum_{j=1}^J m_{ij}$

O valor de a corresponde ao número de pares em que há concordância entre os dois *clusterings*:

$$a = \sum_{i=1}^I \sum_{j=1}^J \binom{m_{ij}}{2} = \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J m_{ij}^2 - \frac{m}{2}$$

O valor de b calcula o número de pares para um dos *clusterings* e retira-lhe o número de pares em que há concordância. Assim ficamos com o número de pares que apenas um dos *clusterings* agrupa.

$$b = \sum_{i=1}^I \binom{m_{i+}}{2} - a = \frac{1}{2} \sum_{i=1}^I m_{i+}^2 - \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J m_{ij}^2 = \frac{1}{2}P - a$$

Em que $m_{i+} = \sum_{j=1}^J m_{ij}$ representa a soma de uma linha da matriz de confusão. A

última expressão envolve um valor P que facilita o cálculo e que será definido mais abaixo.

De forma análoga calculamos c para o outro clustering como sendo o número de pares que apenas o outro *clustering* agrupa.

$$c = \sum_{i=1}^J \binom{m_{+j}}{2} - a = \frac{1}{2} \sum_{i=1}^I m_{+j}^2 - \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J m_{ij}^2 = \frac{1}{2} Q - a$$

Em que $m_{+j} = \sum_{i=1}^I m_{ij}$ representa a soma de uma coluna da matriz de confusão.

De igual forma o valor Q será definido mais abaixo.

E finalmente d é calculado como os restantes pares que não são agrupados por nenhum dos *clusterings* que estão a ser comparados.

$$\begin{aligned} d &= \binom{m}{2} - (a + b + c) = \frac{1}{2} m^2 - \frac{1}{2} \left(\sum_{i=1}^I m_{i+}^2 + \sum_{j=1}^J m_{+j}^2 \right) + \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J m_{ij}^2 \\ &= M - \frac{1}{2} (P + Q) + a \end{aligned}$$

$$P = \sum_{i=1}^I m_{i+}^2 - m \quad \text{e} \quad Q = \sum_{j=1}^J m_{+j}^2 - m \quad (\text{ver Jain e Dubes (1998)})$$

Tabela 5.7. – Exemplo de resultados obtidos a partir das expressões matemáticas desenvolvidas por Jain e Dubes (1988) e Albatineh *et al.* (2006)

m	5000
a	2681025
P	5998904
b	318427
Q	8943330
c	1790640
M	12497500
d	7707408

Para calcular os valores P e Q apresentados na tabela 5.7., usámos a matriz apresentada na tabela 5.5.. Mais concretamente, no caso do valor P , calculámos a soma ao quadrado ($\sum^2$) das linhas da matriz de confusão ao qual foi subtraído o número total

da amostra (m). Em relação ao Q , o método é semelhante ao cálculo do valor P , mas neste caso é necessário calcular a soma ao quadrado ($\sum^2$) a partir das colunas (e não das linhas) da matriz de confusão (ver tabela 5.5.). Os valores a , b , c , d são calculados a partir da matriz apresentada na tabela 5.8..

A partir de uma matriz de concordância, iremos calcular a similaridade entre dois *clusterings*. Neste trabalho, optamos pela medida R , também denominado por *Simple Matching* (SM) (Sokal e Michener 1958; Rand 1971). No Anexo C apresentamos uma lista de medidas de similaridade alternativas. A escolha da medida R baseia-se nas conclusões de um estudo feito por Albatineh *et. al.* (2006).

$$R = \frac{a + d}{a + b + c + d}$$

O valor de R é tanto maior quanto maior for a concordância entre os dois *clusterings*, uma vez que os valores a e d contabilizam os números de pares que ambos os *clusterings* concordam em agrupar e em separar, respectivamente. Se a concordância for total, $b=c=0$, e o índice R toma o valor 1, que é seu valor máximo. O valor 0 expressa a não similaridade, e só acontece se os dois *clusterings* não concordarem em relação ao agrupamento de nenhum dos pares de *clusterings*.

No exemplo que temos seguido esse valor é de aproximadamente 0,831. O resultado do índice R significa neste caso que os dois *clusterings* têm um elevado grau de similaridade. Assim sendo, um deles poderá provavelmente ser descartado.

A partir das tabelas de concordância obtidas para cada par de *clusterings*, e dos respectivos valores de R , construímos uma matriz de similaridade entre todos os pares de *clusterings* em análise (ver tabela 5.8.).

Tabela 5.8. – Exemplo de uma matriz de similaridade

$M_{i,j}^2$	1	2	3	4	5
1	484	36	36	4	4528384
2	100	1	0	1	196
3	0	120409	0	0	0
4	8281	49	9	729	3600
5	78961	1225	100	5476	69169
6	98596	225	81	2304	15129
7	184041	0	0	25	256
8	1089	4	9	100	576
9	0	0	6241	0	0
10	25	1	1	241081	16

Agora podemos passar ao processo de selecção de *clusterings* propriamente dito. Tal como no método anterior, procedemos em duas etapas para reduzir o esforço de análise do utilizador/analista:

- (i) Comparação de *clusterings* dois a dois para os pares com o mesmo número de clusters, com base na matriz de similaridade;
- (ii) Comparação de *clusterings* dois a dois para os pares com diferentes números de clusters.

Na primeira etapa, analisámos as similaridades dos *clusterings* para os pares com o mesmo número de clusters (NC). Podemos ver um exemplo na tabela 5.9. para NC=3.

Tabela 5.9. – Resultados com o número de clusters de três (NC=3)

		Clusterings com o número de 3 (NC=3)							
Clusterings com NC=3	R	A	B	C	D	E	F	G	H
	A	1,00	1,00	0,55	0,99	0,96	0,45	0,65	
	B	1,00		1,00	0,55	0,99	0,96	0,45	0,65
	C	1,00	1,00		0,55	0,99	0,96	0,45	0,65
	D	0,55	0,55	0,55		0,55	0,55	0,49	0,50
	E	0,99	0,99	0,99	0,55		0,97	0,45	0,64
	F	0,96	0,96	0,96	0,55	0,97		0,47	0,63
	G	0,45	0,45	0,45	0,49	0,45	0,47		0,47
	H	0,65	0,65	0,65	0,50	0,64	0,63	0,47	

Neste exemplo, os *clusterings* *A* e *B* são similares, enquanto que o *H* e *D* são muito diferentes. Dada uma matriz de similaridade *S*, e um número máximo *K* de *clusterings* a seleccionar, a estratégia de selecção é a seguinte:

1. Inicialmente o conjunto de *clusterings* interessantes, *CI*, é vazio.
2. Enquanto $|CI| < K$ e houver valores de *S* por analisar, Fazer
 - a. Seja S_{ij} o menor valor (ainda não analisado) de *S*;
 - b. Se *clusterings* c_i e c_j são *diferentes* de todos os que estão em *CI* Então
 - i. *CI* passa a incluir c_i e c_j
3. Apresentar *CI*.

O critério para considerarmos dois *clusterings* como suficientemente diferentes pode ser estabelecido pelo analista.

A estratégia apresentada acima permitiu-nos seleccionar os *clusterings* *A*, *C*, *D* e *H* com um número de clusters (NC) de três (ver tabela 5.10.). O procedimento é análogo em relação ao NC de cinco e NC de dez.

Com os resultados obtidos na primeira etapa, efectuamos a comparação dos *clusterings* seleccionados dois a dois para os pares com diferentes números de clusters, isto é, 5x3 (ver tabela 5.11.), 10x3 e 10x5.

No caso ilustrado na tabela 5.10., foram seleccionados os *clusterings* *A*, *C*, e *K* usando a estratégia desenvolvida para a selecção dos *clusterings*.

Tabela 5.10. – Exemplo de selecção dos *clusterings* com um diferente número de clusters (5x3)

Clusterings (5x3)				
R	A	C	D	H
I	1,00	0,45	0,55	0,65
J	0,55	0,50	1,00	0,73
K	0,36	0,36	0,48	0,43
L	0,38	0,60	0,48	0,53

Recomendamos a consulta do Anexo D para mais detalhes sobre o processo de selecção efectuado.

5.5. Sumário

Neste capítulo apresentámos duas metodologias para a selecção de segmentações (*clusterings*): a metodologia de selecção simples (1ª abordagem) e a metodologia usando o índices de similaridade para a selecção de segmentações (2ª abordagem).

Na nossa opinião, a selecção de segmentações “simples” é vantajosa em relação à segunda abordagem (metodologia), pelo facto de ser de ser mais flexível no que diz respeito à escolha de critérios na selecção de *clusterings*. Ou seja, seria neste caso possível usar outros critérios para além dos critérios propostos, isto é, o NCP e o MaxM. Assim sendo, a metodologia simples é mais adaptada e responde melhor às necessidades dos analistas / peritos.

Em relação à metodologia de selecção com o índice de similaridade R baseia-se em expressões matemáticas e assim permite que profissionais / gestores, sem grandes conhecimentos (ou menos experientes) do negócio, de utilizar esta metodologia e conseguir obter resultados interessantes. Mas tem a desvantagem em relação ao primeiro método relativamente à necessidade de conhecer a medida R para poder usar essa metodologia.

Os resultados obtidos, apresentados no capítulo VI, com ambas metodologia parecem-nos consistentes (ver Anexo D). Entretanto, podemos salientar que não foram observados ganhos significativos de tempo com o uso da metodologia de selecção dita simples em relação à metodologia com a utilização de índice de similaridades, apesar da segunda abordagem necessitar de mais cálculos do que a primeira, como se pode ver na figura 5.2.. Isso porque os resultados são obtidos de modo (semi-) automáticos.

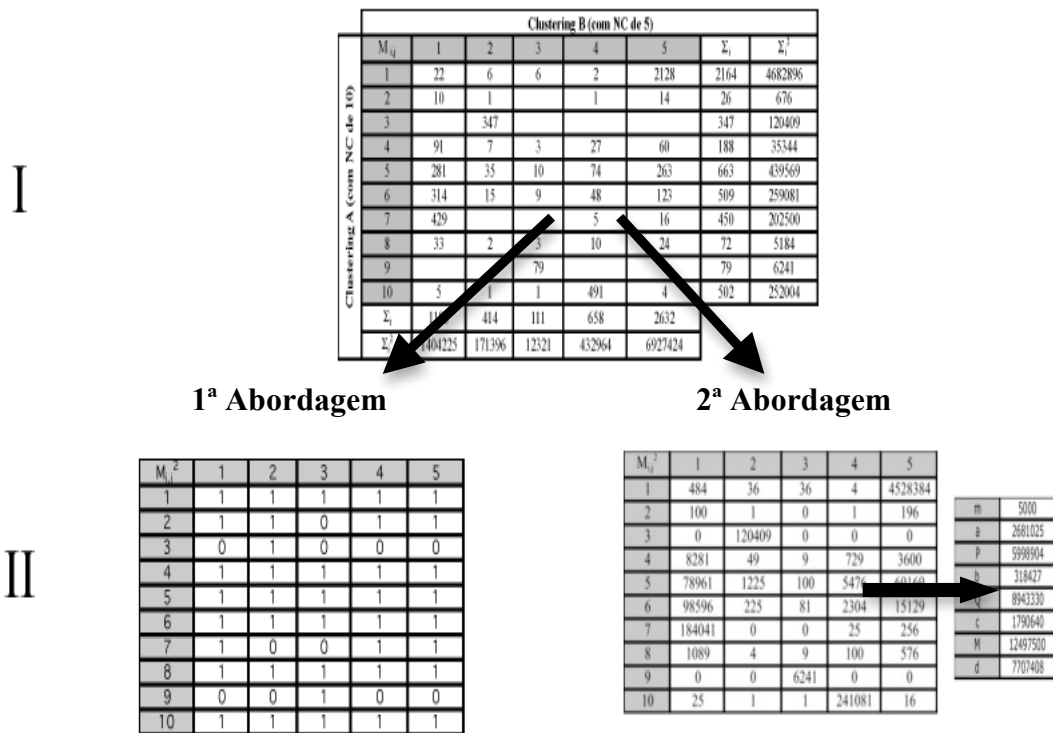


Figura 5.2. – Metodologias de selecção de *clusterings*

No entanto, em ambas as abordagens, é necessária uma análise posterior “manual” das segmentações geradas no fim de analisar e verificar os resultados obtidos (apresentados no próximo capítulo).

CAPÍTULO VI

*“Attempt the impossible in order
to improve your work.”*

(Bette Davis)

Resultados: Perspectiva Técnica

No capítulo V, descrevemos métodos que, dado um grande conjunto de *clusterings* gerados automaticamente, permitem seleccionar um subconjunto constituído por *clusterings* que sejam diferentes uns dos outros.

Estes métodos foram aplicados a um conjunto de *clusterings* gerados com base em dados de consumo de um conjunto de indivíduos. Aqui, são apresentados os dados que foram usados e as transformações realizadas nos mesmos (Secção 6.1). A seguir, vamos descrever as experiências realizadas, identificando os métodos de *clustering* usados e respectivos parâmetros (Secção 6.2). Na Secção 6.3, vamos apresentar as segmentações seleccionadas usando os métodos propostos. Em seguida, é apresentado a descrição e análise dos resultados da investigação realizada (Secção 6.4). Aqui, faremos uma análise numa perspectiva técnica, isto é, relativamente à capacidade de identificar *clusterings* que representam perspectivas diferentes dos dados. No próximo capítulo, vamos analisar os resultados numa perspectiva de aplicação prática. Finalmente será apresentado um sumário da análise realizada (Secção 6.5).

6.1. Preparação dos Dados

Os dados representam despesas de consumo, baseada em informação histórica de um conjunto de pessoas, agregadas por categoria. Ao analisarmos esse conjunto de pessoas, verifica-se que esses padrões/estilos de vida não são infinitos, pelo que será possível classifica-los. A análise da informação contida pode ser considerado como base na previsão de comportamento futuro, descritos pelas seguintes variáveis:

A) Variáveis psico-demográficas (Total de gastos e Idade):

A1) Escala de idade (Esca_Idade): foi obtida com base na distribuição da idade dos clientes de forma a que para cada valor exista um número razoável de clientes (ver tabela 6.1.):

Tabela 6.1. – Descrição da escala de idade

Valor	Intervalo de Idades
1	25 anos ou menos
2	dos 26 aos 30 anos
3	dos 31 aos 40 anos
4	dos 40 aos 55 anos
5	55 anos ou mais

A2) Escala de gastos (Esca_Gastos): foi obtida por comparação do valor total de gastos feito pelo cliente com os percentis obtidos com base em todos os clientes. O valor é estabelecido separadamente para os clientes de cada valor da escala de idade da seguinte forma (ver tabela 6.2.):

Tabela 6.2. – Descrição da variável escala de gastos

Valor	Percentil
1	inferior a 15%
2	entre 15% e 35%
3	entre 35% e 65%
4	entre 65% e 87%
5	superior a 87%

B) Variáveis de consumo (Como e onde gasta):

B) 10 variáveis representando os gastos: automóveis (CARS), viagens (TRAVEL), vestuário (CLOTHES), lar (HOME), serviços de beleza

(BEAUTY), alimentação (FOOD), crianças (KIDS), educação e cultura (EDUCULTURE), lazer (HOBBIESFUN), e levantamentos de dinheiro (MONEY).

As variáveis originais relativas a gastos representam o valor total gasto em cada categoria. Com este tipo de variáveis é possível obter segmentos que representam os perfis em termos dos volume de gastos. Parece-nos igualmente importante representar o peso de cada categoria de despesas na despesa total feita pelo cliente. As tabelas 6.3. e 6.4. mostram dois clientes com gastos em termos de volumes e em termos de proporção respectivamente. Como se pode ver, estes dois clientes são muito diferentes de acordo com a primeira representação e muito semelhantes de acordo com a segunda. Desta forma, vamos obter segmentações que nos dão perspectivas diferentes sobre o mesmo conjunto de clientes.

Tabela 6.3. – Exemplo de resultados em termos dos volumes de gastos

Variáveis em termos dos volume de gastos										
Clientes	CARS	TRAVEL	CLOTHES	HOME	BEAUTY	FOOD	KIDS	EDUCULTURE	HOBBIESFUN	Total
1	808,7	1772,5	931,3	3578,5	7468,3	5614,5	460,8	872,9	1912,4	23419,9
2	277,0	588,0	398,7	1307,0	2046,8	2120,8	179,6	383,3	720,4	8021,4

Tabela 6.4. – Exemplo de resultados em termos de proporção

Variáveis em termos de proporção de gastos									
Clientes	pCARS	pTRAVEL	pCLOTHES	pHOME	pBEAUTY	pFOOD	pKIDS	pEDUCULTURE	pHOBBIESFUN
1	3,5%	7,6%	4,0%	15,3%	31,9%	24,0%	2,0%	3,7%	8,2%
2	3,5%	7,3%	5,0%	16,3%	25,5%	26,4%	2,2%	4,8%	9,0%

Assim, criámos um novo conjunto de 10 variáveis representando a proporção do valor de cada uma das 10 variáveis de gastos originais no total dos gastos do cliente (e.g. X% dos gastos do cliente são em roupa, Y% em carros, etc...). Uma segmentação obtida com base nestas variáveis identifica perfis de clientes em termos da forma como distribuem as suas despesas nas várias categorias.

Ao analisar os dados foi identificado um cliente com todos os gastos a zero.

6.2. Descrição das Experiências: Métodos e Parâmetros

A metodologia de selecção, apresentada no capítulo V, foi aplicada em 24 segmentações obtidas usando:

- 4 métodos de clustering: k-médias (K) e hierárquico com funções de distância centróide (C), *average linkage* (H) e *Ward* (W);
- 3 valores diferentes do número de clusters (NC): 3, 5 e 10;
- 2 conjuntos de variáveis: escalas + gastos em valor (F) e escalas + gastos em proporção (P).

A seguir, vamos explicar como determinamos os clusters.

6.2.1. Determinação dos Clusters

Esta secção tem por objectivo explicar como foi feita a respectiva atribuição dos clusters e as ferramentas usadas.

Para determinar os clusters recorremos ao software *Enterprise Guide* 4 da SAS. O *SAS Enterprise Guide* é uma aplicação para fazer exploração de dados que consegue lidar com grandes quantidades de dados. Esta aplicação permite, por exemplo, aceder e manipular os dados, executar código SAS, e gerar resultados de forma conveniente.

A interface (gráfica) é do tipo *point-and-click*²⁸ (ver figura 6.1.), sendo que a grande vantagem de utilizar a interface do *Enterprise Guide* é que as operações podem ser executada com uma grande rapidez sem que seja necessário um conhecimento profundo da linguagem de programação SAS.

²⁸ O “apontar e clicar” (*Point-and-click*) é a acção de um utilizador de computador de deslocar o cursor para um determinado local num ecrã (ponto) e em seguida, premiar um botão do rato, normalmente o esquerdo (clique).

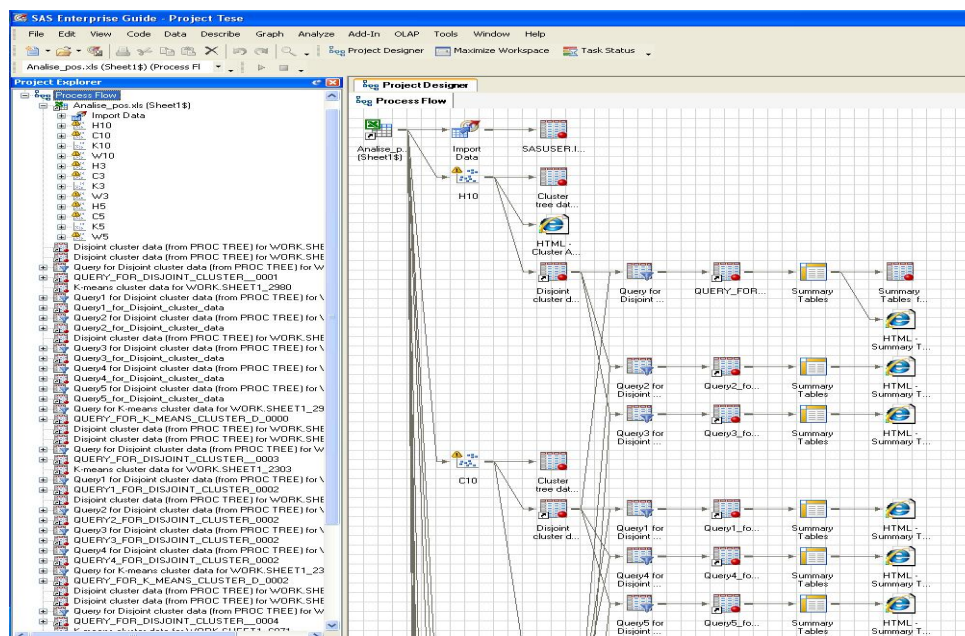


Figura 6.1. – Interface do *SAS Enterprise Guide 4.2*

Usamos os seguintes algoritmos k-médias (K), centróide (C), average linkage (H) e *Ward* (W) disponíveis com a aplicação *SAS Enterprise Guide*, ilustrado na figura 6.2.),

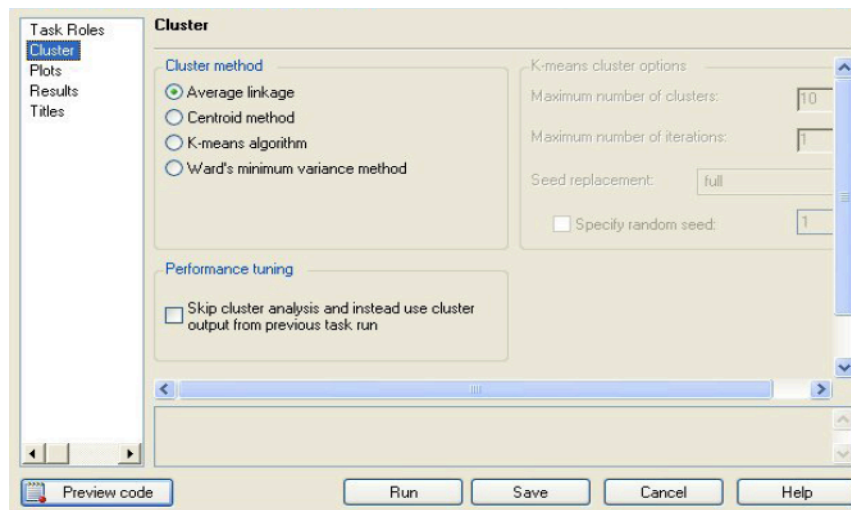


Figura 6.2. – Métodos de clustering disponíveis no *SAS Enterprise Guide*

Os parâmetros usados foram os que estão pré-definidos no *SAS Enterprise Guide*.

A figura 6.3, ilustra um exemplo de determinação de clusters usando o algoritmo k-médias com dez clusters (Número de Cluster =10).

	Esca- Idade	Esc- Gasto	CARS	TRAVEL	CLOTHES	HOME	BEAUTY	FOOD	KIDS	EDU CULT URE	HOBBIES FUN	MONEY	CLUSTER	DISTANCE
1	5		720.82	300	118.8	1375.02	281.7	4893.49	60.7	171.3	951.9	81.02	7	3078.0060758
2	2		0	0	0	75.21	0	57.06	0	0	56.65	402.13	4	719.65248691
3	3		0	657.06	0	0	0	401.75	0	0	0	0	4	953.71607085
4	5		0	4093.5	506.7	176.71	534.7	333.05	54.9	21.6	213.55	0	4	4083.5452018
5	5		488.03	37.9	485.15	947.11	446.79	7302.96	38.05	0	721	172.08	7	5321.2235788
6	5		1765.79	0	533.45	2002.74	779.08	2702.38	18.3	75.59	207.15	293.18	7	1701.3577349
7	5		1539.44	47.2	2570.35	1925.81	2429.14	2629.86	292.75	500.66	631.48	403.11	7	3161.1528126
8	5		1549.94	45.55	1430	635.94	419.11	4207.17	0	0	272.43	0	7	2453.7554493
9	5		666.95	0	254.4	492.63	2626.73	981.26	0	0	85	978.03	7	2458.3949513
10	5		2012.54	22.7	188.99	3566.78	398.46	2040.73	311.8	64.77	100.68	443.1	7	3113.7168582
11	5		1286	0	1019.4	150.21	912.51	8139.42	20	0	806.14	324.56	7	6189.6957548
12	4		215.49	0	27.5	703.57	61.55	1724.36	0	0	329.44	0	7	1222.5455664
13	4		11.25	0	2154.15	531.97	320.45	1345.87	70.75	299	113.28	147	7	2096.5603777
14	5		1998.12	157.85	969.13	768.64	948.07	3452.25	0	92.85	564.42	698	7	1930.7139331
15	4		26.27	0	49.9	155	2398.68	399.77	0	0	0	162.64	4	2215.7361132
16	4		742.79	0	54.6	114.45	207.71	1600.09	0	322.7	35.18	71.49	7	1112.061253
17	2		530.59	0	0	0	0	14.44	0	0	0	0	4	1024.6608179
18	5		961.96	4002	1793.45	157.43	2592.25	3125.28	0	65	79.27	83.85	7	4483.8739519
19	4		0	0	1895.4	968.72	476.02	698.62	0	24.5	90	80.5	4	1838.0800497
20	2		0	0	0	0	0	0	0	0	0	209.51	4	851.2416548
21	5		2314.09	0	675.48	1285.27	583.41	4000.65	86.95	0	148.34	822.37	7	2549.5227133
22	4		46	0	167.78	324.62	436.56	2663.17	59.65	0	27	91.4	7	1323.9794533
23	4		225.91	0	214.81	514.17	1066.54	814.31	0	26.09	76.9	132.5	4	981.16452374
24	4		58.5	501.91	74.8	242.82	408.81	1862.33	0	169.8	181.69	270	7	1177.3713245
25	3		0	0	0	59.9	66.98	996.98	0	0	0	0	4	796.50248142
26	4		853.92	0	217.4	27.73	15.63	4028.2	0	0	316.69	25	7	2253.353679
27	5		664.5	2705.38	2803.55	148.6	2724.54	1272.96	100.4	155.96	271.63	271.3	7	4000.2042085
28	4		31.84	0	0	296.18	622.09	2266.95	0	25.95	1265.35	30.65	7	1640.7948277
29	3		68.5	0	29.95	29.59	117.06	1131.22	0	0	15.2	0	4	853.68479855
30	3		0	0	0	123.51	0	1793.19	0	79.21	123.6	0	4	1372.2388815
31	2		0	0	0	0	630.39	0	0	0	0	0	4	999.94402975
32	5		1165.35	843.08	1585	1333.62	717.98	3976.67	0	0	95.99	17.75	7	2358.6526952
33	5		970.06	1284.19	2090.78	819.79	3368.79	6428.77	320.43	771.51	1410.02	345.85	10	4411.5965674
34	4		376.25	0	344.5	355.01	1351.49	1608.48	0	14.85	15.6	122.27	7	1180.3854716
35	3		0	0	0	163.68	55	476.7	44.9	0	0	124.07	4	587.61327743
36	2		26.27	0	30.96	76.78	0	89.1	0	0	0	0	4	868.64636936
37	5		2088.71	36	722.74	1615.34	473.48	4477.39	0	72	209.29	335.58	7	2855.8435122
38	3		84.32	0	0	136.13	0	820.14	0	0	0	34.25	4	690.76255953

Figura 6.3. – Exemplo de determinação de clusters com recurso a ferramenta *SAS Enterprise Guide*

A coluna *CLUSTER* (ver figura 6.3.) mostra a respectiva atribuição dos clusters dos elementos da primeira coluna. Por exemplo, o primeira elemento está incluído no cluster sete, enquanto o elemento trinta e três está no cluster dez.

Convém referir que o software da *SAS* é uma ferramenta muito utilizada pelos bancos em Portugal por exemplo.

6.3. Selecção de Segmentações

A metodologia proposta permitiu seleccionar um pequeno conjunto que foi posteriormente analisado de forma manual. Com as metodologias desenvolvidas no capítulo V, foi possível obter os seguintes resultados para cada um dos métodos (abordagens), apresentados na tabela 6.5..

Tabela 6.5. – Resultados obtidos usando as metodologias desenvolvidas

	(Pré-) selecção		Seleccção Final
	1ª Abordagem	2ª Abordagem	
5*3	<u>H5P</u> , K3P , W3F	K3P , <u>K5P</u> , H3F	K3P , H5P, K5P
10*3	<u>K10P</u> , K3P	H3F, <u>W3F</u> , W10P	W3F, K10P
10*5	K10P, <u>K5P</u> , W10P	<u>W5F</u> , W10P	K5P, W5F, W10P

Legendas: valores sublinhados – melhores resultados (das segmentações)

valores a cores – resultados comuns em ambas as abordagens

De um modo geral, conseguimos o nosso objectivo de obter segmentações diferentes usando as metodologias desenvolvidas. Na primeira abordagem (selecção dita simples), é interessante constatar na pré-selecção das segmentações (ver tabela 6.5.) que os resultados obtidos na selecção das segmentações (*clusterings*) foram variáveis em termos de gastos em proporção (P), destacando-se o método de clustering k-médias (K), enquanto que na segunda abordagem (usando índices de similaridades) os resultados foram mais “diversificados”.

Como podemos constatar na tabela 6.5., em ambas as abordagens foi pré-seleccionadas seis segmentações, *i.e.*, três segmentações com três clusters (NC=3); duas com cinco; e uma segmentação com dez clusters na primeira abordagem. Para a segunda abordagem foi pré-seleccionadas duas segmentações com três clusters; duas com cinco clusters; e duas segmentações com 10 clusters.

No entanto, para a selecção final escolhemos sete segmentações ao seleccionar, em primeiro lugar, os resultados comuns e em segundo lugar, ao escolher os melhores resultados obtidos em relação à cada uma das abordagens.

As segmentações escolhidas são:

- (i) a segmentação W3F - clustering hierárquico, usando a distância de *Ward*, com 3 clusters obtido com base nas variáveis escalas + gastos em valor (F);
- (ii) a K3P - clustering não-hierárquico, usando o k-médias, com 3 clusters obtido com base nas variáveis escalas + gastos em proporções (P);

- (iii) a K5P - clustering não-hierárquico, usando o k-médias, com 5 clusters obtido com base nas variáveis escalas + gastos em proporções (P);
- (iv) a H5P - clustering hierárquico, usando a distância de *average linkage*, com 5 clusters obtido com base nas variáveis escalas gastos em proporções (P);
- (v) a W5F - clustering hierárquico, usando a distância de *Ward*, com 5 clusters obtido com base nas variáveis escalas + gastos em valor (F);
- (vi) a K10P - clustering não-hierárquico, usando o k-médias, com 5 clusters obtido com base nas variáveis escalas + gastos em proporções (P);
- (vii) e a segmentação W10P - clustering hierárquico, usando a distância de *Ward*, com 10 clusters obtido com base nas variáveis escalas + gastos em proporções (P).

Para mais detalhes no processo de selecção dos clusterings, recomendamos a consulta do Anexo D.

A seguir, vamos descrever e analisar os resultados de cada uma das segmentações escolhidas.

6.4. Descrição e Análise dos Resultados (das Segmentações)

Antes de analisar as segmentações realizadas é importante ter uma ideia de qual é o perfil médio do cliente. As tabelas que se seguem apresentam os valores médios das variáveis para os clientes da amostra disponibilizada.

Tabela 6.6. – Valor médio da escala de gastos e idade da amostra

Escala de Gastos	Escala de Idade
4,2	3,4

O valor médio da escala de idades indica que a amostra usada é constituída por clientes não muito jovens (o valor 4 representa o segmento entre os 40 e os 55 anos), como se pode ver na figura 6.4.. O valor médio da escala de gastos é superior ao valor esperado dada a forma como a variável é calculada, e que seria de 3. Isto indica que a

amostra pode não ser completamente aleatória.

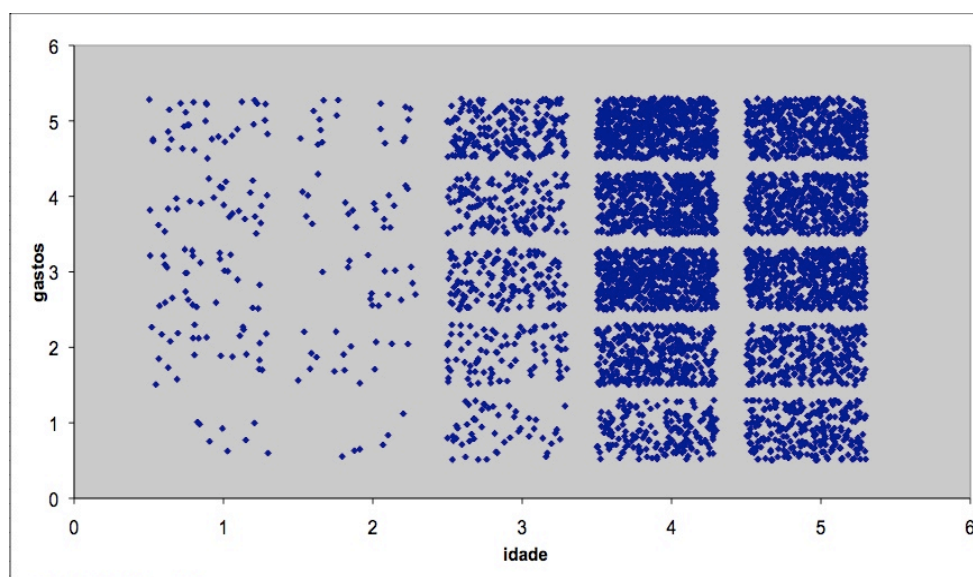


Figura 6.4. – Representação gráfica da escala de gastos e idade da amostra

De acordo com os seus valores médios, podemos separar as variáveis de gastos em 3 grupos: valores baixos (TRAVEL, KIDS, EDUCULTURE e HOBBIESFUN, com pesos nas despesas inferiores a 5%), valores médios (CARS, CLOTHES, HOME, BEAUTY e MONEY, com valores entre 8% e 15%) e a variável FOOD com um peso muito importante nas despesas efectuadas (ver tabela 6.7.).

Tabela 6.7. – Valores médios e peso nas despesas efectuadas

	CARS	TRAVEL	CLOTHES	HOME	BEAUTY	FOOD	KIDS	EDUCULTURE	HOBBIESFUN	MONEY
Valor	549,1	229,4	477,6	531,5	492,9	1363,7	49,9	102,3	178,5	414,1
Prop.	14%	4%	10%	11%	12%	34%	1%	2%	3%	8%

Segue-se a análise de cada uma das segmentações seleccionadas.

6.4.1. Segmentação W3F

A segmentação W3F (ver tabela 6.8.) contém três clusters obtidos com o método de clustering hierárquico, usando a distância de *Ward*, com base nas variáveis escalas e gastos em valor (F).

Tabela 6.8. – Segmentação W3F

Segmentação W3F													
Cluster	Nr. Clientes	Escala de Idade	Escala de Gastos	CARS	TRAVEL	CLOTHES	HOME	BEAUTY	FOOD	KIDS	EDUCULTURE	HOBBIESFUN	MONEY
1	3300	4,2	2,7	276,6	77,0	188,0	198,7	207,0	468,3	14,5	60,8	72,9	112,4
2	1699	4,1	4,7	974,8	525,7	1040,4	1178,4	1048,6	3103,5	118,8	183,0	383,6	1000,2
3	1	4,0	5,0	176814,9	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Esta segmentação tem, na realidade, apenas 2 segmentos que são o primeiro e segundo cluster, enquanto que um dos clusters (3) tem apenas um cliente, com gastos muito elevados e apenas na categoria automóvel (CARS). De facto, podemos considerar este cliente como sendo um *outlier*, pelo facto de ter gastos em apenas uma categoria.

Apesar de as segmentações terem sido obtidas com os gastos representados em valor absoluto, analisámos os perfis correspondentes em termos da sua distribuição, como se pode ver na tabela 6.9..

Tabela 6.9. – Segmentação W3F (em proporção)

Segmentação W3F (em proporção)											
Cluster	Nr. Clientes	pCARS	pTRAVEL	pCLOTHES	pHOME	pBEAUTY	pFOOD	pKIDS	pEDUCULTURE	pHOBBIESFUN	pMONEY
1	3300	16%	4%	11%	11%	13%	32%	1%	3%	3%	7%
2	1699	11%	4%	10%	12%	11%	36%	1%	2%	3%	10%
3	1	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Os dois restantes clusters têm perfis de distribuição das despesas pelas várias categorias muito semelhantes (gastos em proporção). A distinção entre ambos faz-se essencialmente na escala de gastos, como se pode ver na tabela 6.8.. Os clientes do cluster 2 são provavelmente clientes com uma situação financeira melhor do que os do cluster 1.

6.4.2. Segmentação K3P

A segmentação K3P (ver tabela 6.10.) contém com 3 clusters obtidos com o método e clustering não-hierárquico, usando o k-médias com base nas variáveis escalas e gastos em proporções (P).

Tabela 6.10. – Segmentação K3P

Segmentação K3P													
Cluster	Nr. Clientes	Escala de Idade	Escala de Gastos	CARS	TRAVEL	CLOTHES	HOME	BEAUTY	FOOD	KIDS	EDUCULTURE	HOBBIESFUN	MONEY
1	1669	4,2	3,3	615,7	288,1	624,3	599,5	667,7	781,0	55,9	123,1	207,4	237,6
2	469	4,2	3,4	446,7	130,9	299,1	445,6	315,2	811,9	26,9	55,5	117,4	1912,4
3	2862	4,1	3,5	527,1	211,3	421,4	506,0	420,2	1793,9	50,2	97,9	171,6	271,4

Podemos constatar que em ambos os cluster o nível de gastos (escalão de Gastos) é muito similar (<65%/Escalão idade), sendo ligeiramente superior no cluster 3.

Esta segmentação distingue os clientes em 3 segmentos com perfis muito claros de distribuição das despesas realizadas, com valores significativamente mais elevados em 1 ou mais categorias, como se pode ver na tabela 6.11..

Tabela 6.11. – Variáveis com a proporção de despesas significativamente acima da média da segmentação K3P

Cluster	Nr. Clientes	Variáveis com proporção de despesa significativamente acima da média
1	1669	CARS 18% CLOTHES 14% HOME 14% BEAUTY 18%
2	469	MONEY 40%
3	2862	FOOD 45%

O cluster 1 contém clientes com uma proporção elevada ($\frac{3}{4}$ dos gastos) das suas despesas em automóveis, vestuário, lar e serviços de beleza. É difícil, sem um conhecimento mais profundo da natureza das variáveis e do negócio, identificar o tipo de clientes a que este perfil corresponde, pelo facto de ser em termos socioeconómicos e, por isso, não o tentaremos fazer..

É interessante notar que cada segmento tem uma distribuição de idades e de gastos muito semelhante à global, o que significa que representam fatias transversais do mercado. Ou seja, estes perfis de consumo parecem representar perfis sócio-psicológicos que não mudam com a idade nem com as posses.

6.4.3. Segmentação K5P

A segmentação K5P (ver tabela 6.12.) contém 5 clusters obtidos com o método de clustering não-hierárquico, usando o k-médias, com base nas variáveis escalas e gastos em proporções (P).

Tabela 6.12. – Segmentação K5P

Segmentação K5P													
Cluster	Nr. Clientes	Escala de Idade	Escala de Gastos	CARS	TRAVEL	CLOTHES	HOME	BEAUTY	FOOD	KIDS	EDUCULTURE	HOBBIESFUN	MONEY
1	1185	4,2	3,2	656,8	290,5	672,3	379,6	768,2	795,2	59,6	98,6	110,6	229,0
2	414	4,2	3,4	454,7	134,8	291,2	347,1	316,4	792,3	21,0	52,5	105,6	2062,3
3	111	4,1	3,5	497,9	141,2	251,7	235,2	290,4	822,3	84,4	114,6	1724,0	263,4
4	658	4,1	3,5	555,4	277,6	495,0	1361,7	418,0	999,0	55,1	159,2	146,4	297,2
5	2632	4,1	3,5	516,1	208,5	424,5	433,9	424,1	1823,5	47,4	97,1	163,4	273,7

Se, como seria de esperar, esta segmentação é semelhante à K3P, há no entanto, algumas diferenças interessantes, como se pode ver na tabela 6.13..

Tabela 6.13. – Variáveis com a proporção de despesa significativamente acima ou abaixo da média da segmentação K5P

Cluster	Nr. Clientes	Variáveis com proporção de despesa significativamente acima (+) ou abaixo (-) da média		
1	1185	CARS 20% (+)	CLOTHES 16% (+)	BEAUTY 22% (+)
2	414	MONEY 43% (+)		
3	111	HOBBIESFUN 34% (+)		
4	658	HOME 33% (+)		
5	2632	FOOD 47% (+)		

O cluster 1 da K3P dividiu-se em clusters mais especializados. O cluster 1 contém agora clientes que têm proporções ainda mais elevadas de despesas em automóveis, vestuário e serviços de beleza mas valores normais nas despesas relacionadas com o lar. Os clientes que contribuíam para o valor elevado deste último tipo de despesas no cluster 1 da K3P estão agora no cluster 4.

Por outro lado, um pequeno grupo de clientes que contribuíam para um valor ligeiramente mais elevado de despesas em lazer no cluster 1 da K3P está agora separado no cluster 3.

6.4.5. Segmentação H5P

A segmentação H5P (ver tabela 6.14.) contém 5 clusters obtidos com o método de clustering hierárquico, usando a distância de *average linkage*, com base nas variáveis escalas e gastos em proporções (P).

Tabela 6.14. – Segmentação H5P

Segmentação H5P													
Cluster	Nr. Clientes	Escala de Idade	Escala de Gastos	CARS	TRAVEL	CLOTHES	HOME	BEAUTY	FOOD	KIDS	EDUCULTURE	HOBBIESFUN	MONEY
1	4687	4,2	3,4	493,2	240,1	500,6	554,9	514,6	1424,9	50,6	104,6	155,1	434,3
2	226	4,2	3,4	1799,7	48,1	118,5	170,8	146,1	404,0	5,4	32,5	57,7	89,5
3	55	4,2	3,4	425,5	139,1	162,3	183,7	269,1	671,2	5,8	50,1	2759,5	236,4
4	21	4,1	3,4	81,3	50,2	289,1	213,1	91,8	230,8	29,0	513,0	23,5	60,2
5	11	4,1	3,4	200,3	224,2	17,2	328,4	257,6	593,8	953,8	35,4	34,2	48,1

Esta segmentação tem a característica de que cada cluster se especializa claramente numa variável distinta de gastos, como se pode ver na tabela 6.15.. E também tem a característica de ter uma escala de gastos muito semelhante para todos os clusters.

Tabela 6.15. – Variáveis com proporção de despesa significativamente acima da média da Segmentação H5P

Cluster	Nr. Clientes	Variáveis com proporção de despesa significativamente acima da média
1	4687	FOOD 32%
2	226	CARS 63%
3	55	HOBBIESFUN 56%
4	21	EDUCULTURE 32%
5	11	KIDS 35%

Esta segmentação caracteriza-se por ter essencialmente um segmento de grande dimensão (cluster 1) e tem a particularidade de basicamente separar 4 mini-/micro-segmentos da população geral, sendo que os cluster 3, 4 e 5 têm gastos em variáveis que normalmente não têm gastos muito significante em relação às outras variáveis.

Neste caso, podemos perguntar se a identificação de nichos seria interessante do ponto de visto do negócio. Por exemplo, o cluster 5 trata-se mais de um micro-segmento pelo facto de conter apenas onze elemento. Neste último caso, podemos ainda perguntar se esse segmento é suficientemente grande para ser accionável.

6.4.5. Segmentação W5F

A segmentação W5F (ver tabela 6.16.) contém 5 clusters obtidos com o método de clustering hierárquico, usando a distância de *Ward*, com base nas variáveis escalas e gastos em valor (F).

Tabela 6.16. – Segmentação W5F

Segmentação W5F													
Cluster	Nr. Clientes	Escala de Idade	Escala de Gastos	CARS	TRAVEL	CLOTHES	HOME	BEAUTY	FOOD	KIDS	EDUCULTURE	HOBBIESFUN	MONEY
1	3300	4,2	2,7	276,6	77,0	188,0	198,7	207,0	468,3	14,5	60,8	72,9	112,4
2	1666	4,1	4,7	964,6	503,4	1053,3	1179,6	1058,1	3046,8	120,8	179,6	383,3	720,4
3	31	4,1	5,0	1130,8	208,7	372,5	831,3	578,5	1565,0	22,2	144,8	367,6	15836,9
4	2	1,0	5,0	7106,0	24049,9	707,3	5506,6	408,4	74237,7	0,0	3646,6	898,0	4130,3
5	1	4,0	5,0	176814,9	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Esta segmentação é, como se podia esperar, muito semelhante à segmentação W3F. A diferença está na divisão do cluster 2 da W3F em três segmentos na W5F. Um destes clusters (4) tem apenas 2 clientes, que também podem ser considerados *outliers*, por serem muito jovens (Escala de Idade 1) e terem gastos muito elevados em viagens e comida.

Constata-se que nesta segmentação existe dois clusters (1 e 2) de grandes dimensões (ver tabela 6.16.), enquanto que nos restantes três clusters tratam-se mais de

mini-/micro- segmentos, mas com a particularidade de ter um valor máximo da escala de gastos (5) .

O cluster 1, é, em um geral, um cluster muito homogéneo, mas destaca-se gastos mas elevado em alimentação em relação às outras variáveis deste cluster.

Tabela 6.17. – Variáveis com valor significativamente acima ou abaixo da média do Cluster 3

Cluster	Nr. Clientes	Variáveis com valor significativamente acima (+) ou abaixo (-) da média				
3	31	TRAVEL 208 (-)	CLOTHES 372 (-)	BEAUTY 578 (-)	FOOD 1565 (-)	KIDS 22 (-)
						MONEY (prop.) 75% (+)

Para além disso, os gastos em viagens, vestuário, serviços de beleza, crianças e, particularmente, em alimentação são comparativamente baixos relativamente ao cluster 2. No entanto, este cluster tem apenas 31 clientes (ver tabela 6.17.), colocando-se a questão sobre se será suficientemente grande para ser “accionáveis” (*actionable*), ou seja, se vale a pena criar uma estratégia de marketing específica ou se, por outro lado, será suficientemente pequeno e com características suficientemente interessantes para ser possível analisar cada cliente individualmente.

E finalmente, o cluster 5 conta com um elemento que pode ser considerado como um *outlier* (100% dos gastos totais foram efectuados em “carros”).

6.4.6. Segmentação K10P

A segmentação K10P (ver tabela 6.18.) contém 5 clusters obtidos com o método de clustering não-hierárquico, usando o k-médias, com base nas variáveis escalas e gastos em proporções (P).

Tabela 6.18. – Segmentação K10P

Segmentação K10P													
Cluster	Nr. Clientes	Escala de Idade	Escala de Gastos	CARS	TRAVEL	CLOTHES	HOME	BEAUTY	FOOD	KIDS	EDUCULTURE	HOBBIESFUN	MONEY
1	2164	4,1	3,5	475,6	177,9	418,2	448,9	458,4	1979,2	42,4	91,1	138,3	302,0
2	26	3,8	3,7	419,6	184,7	194,7	271,8	320,2	990,7	1176,9	121,3	118,6	157,8
3	347	4,2	3,3	341,2	78,7	243,6	341,1	305,3	773,0	13,1	46,4	104,8	2254,4
4	188	4,2	3,7	530,3	2023,4	495,2	540,2	520,7	1176,7	53,8	140,3	181,6	309,3
5	663	4,1	3,2	1362,1	112,5	210,6	345,7	259,1	760,9	24,7	56,3	214,1	192,4
6	509	4,2	3,4	285,8	171,1	1383,0	489,0	597,4	954,9	77,1	106,5	120,1	300,6
7	450	4,4	3,1	290,8	131,0	456,6	437,4	1168,7	838,0	54,1	80,5	119,9	228,8
8	72	3,9	3,1	360,9	145,8	320,0	382,9	254,8	679,9	16,3	915,5	100,4	207,1
9	79	4,2	3,5	461,3	138,7	211,4	218,5	281,1	791,8	7,0	58,1	2239,6	210,7
10	502	4,1	3,5	489,4	214,5	421,8	1473,0	434,7	1077,6	61,5	140,6	156,0	313,1

Esta segmentação tem a característica de que cada cluster se especializa claramente numa variável distinta de gastos, como se pode ver na tabela 6.19..

Tabela 6.19. – Grau de especialização de cada cluster da segmentação K10P

	CARS	TRAVEL	CLOTHES	HOME	BEAUTY	FOOD	KIDS	EDUCULTURE	HOBBIESFUN	MONEY
Cluster	5	4	6	10	7	1	2	8	9	3
Nr. Cli.	663	188	509	502	450	2164	26	72	79	347
Prop.	45%	43%	36%	36%	41%	50%	30%	35%	42%	47%

Constata-se que, com excepção do cluster 2, associado à variável de despesas com crianças e que tem apenas 26 clientes, todos os clusters têm uma dimensão razoável. É também interessante observar que os valores médios das escalas de idade e de gastos são semelhantes para todos os clusters (ver figura 6.5.). Em particular, os segmentos são muito semelhantes relativamente ao perfil etário, como se pode ver na figura abaixo. A excepção é o cluster 2, que se concentra muito nos escalões 3 e 4. Serão, por ventura, os jovens pais. Note-se, no entanto, que este é o cluster mais pequeno (26 elementos), logo mais sujeito a variância. Por isso, a diferença observada pode ser um acaso.

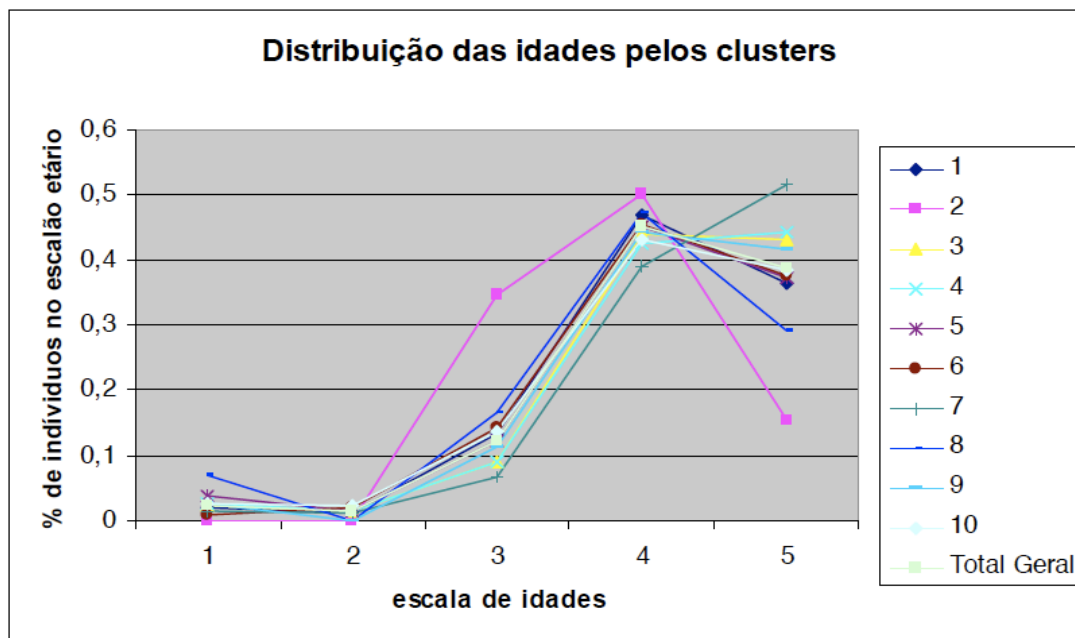


Figure 6.5. – Distribuição das idades pelos clusters do K10P

6.4.7. Segmentação W10P

A segmentação W10P (ver tabela 6.20.) contém 10 clusters obtidos com o método de clustering hierárquico, usando a distância de *Ward*, com base nas variáveis escalas e gastos em proporções (P).

Tabela 6.20. – Segmentação W10P

Segmentação W10P													
Cluster	Nr. Clientes	Escala de Idade	Escala de Gastos	CARS	TRAVEL	CLOTHES	HOME	BEAUTY	FOOD	KIDS	EDUCULTURE	HOBBIESFUN	MONEY
1	147	4,5	2,8	195,8	116,9	269,5	364,4	1280,9	453,0	15,2	29,0	67,3	198,3
2	476	4,2	2,8	231,9	168,8	183,2	212,1	167,9	1558,4	9,7	54,3	50,9	105,1
3	431	4,1	3,4	531,7	246,6	356,5	1384,4	340,3	969,0	39,4	91,6	142,1	321,5
4	251	4,2	3,3	334,5	59,1	204,0	294,3	225,8	754,7	11,6	47,4	71,3	2660,1
5	154	4,3	3,6	457,7	1786,8	479,2	385,0	424,7	798,2	36,2	98,0	141,1	361,8
6	566	4,1	3,2	1432,7	80,4	221,8	318,6	253,0	771,1	18,2	54,5	217,9	212,9
7	259	4,2	3,1	230,1	111,3	1468,0	538,6	508,8	714,1	82,3	105,4	118,8	285,9
8	39	4,3	3,5	504,1	179,8	199,7	235,1	282,6	753,7	6,9	70,7	3606,1	287,1
9	1489	4,2	3,6	434,3	289,3	662,7	586,2	740,2	1505,4	90,1	163,5	199,2	377,0
10	1188	4,1	3,7	577,5	118,9	406,1	480,9	454,2	2009,8	43,2	93,2	139,8	304,9

Esta segmentação apresenta um elevado grau de especialização de cada cluster, à semelhança da segmentação anterior, K10P. No entanto, como se pode ver na tabela 6.21., a especialização é geralmente mais pronunciada, em particular no caso das categorias vestuário, serviços de beleza, alimentação, lazer e levantamentos de dinheiro. A segmentação W10P fornece informação complementar em relação a K10P, e assim sendo a questão de redundância não se aplica por ser efectivamente diferentes.

No entanto, podemos observar que na W10P existe dois grandes grupos (cluster 9 e 10), enquanto na K10P existe apenas um grande grupo.

Tabela 6.21. – Grau de especialização de cada cluster da segmentação W10P

	CARS	TRAVEL	CLOTHES	HOME	BEAUTY	FOOD	KIDS	EDUCULTURE	HOBBIESFUN	MONEY
Cluster	6	5	7	3	1	2 / 9 / 10			8	4
Nr. Clientes	566	154	259	431	147	476 / 1489 / 1188			39	251
Proporção	48%	46%	43%	37%	56%	72% / 30% / 45%			52%	54%

Podemos observar ainda que a distribuição dos clientes pelos clusters é mais homogénea relativamente à segmentação K10P.

O cluster 9 é o maior cluster, com 1489 clientes e que não se destaca particularmente em nenhuma categoria, com excepção dos gastos em vestuário e serviços de beleza.

É também interessante notar que nesta segmentação o valor médio dos valores da escala de gastos varia entre clusters. Em particular, os clusters 1 e 2 apresentam valores significativamente mais baixos enquanto que os clusters 5, 9 e 10 apresentam valores mais elevados. Isto permite distinguir melhor os clusters 2 e 10 que, embora concentrem as suas despesas em comida, fazem-no em níveis diferentes não só em termos da proporção dos gastos mas também do seu valor. Permite também distinguir melhor os clusters 1 e 9 no que diz respeito às despesas em serviços de beleza.

6.5. Sumário

Neste capítulo foram aplicadas as metodologias descritas no capítulo V a dados com despesas de um conjunto de pessoas.

De um modo geral, concluímos que algumas segmentações, em termos de resultado, são muito semelhantes. Por exemplo, no caso das segmentações W3F e W5F que foram obtidos usando o mesmo método de *clustering*, mas com um número de clusters diferente. Outra conclusão geral é que algumas segmentações com mais segmentos podem ser consideradas como especializações de outras com menos segmentos, dividindo em sub-segmentos mais especializados, que, nalguns casos, são mini- ou micro-segmentos, como por exemplo no caso da segmentação H5P. Estas especializações podem ser interessantes para o negócio, porque oferecem um maior potencial de vantagem competitiva ao procurar satisfazer às exigências de um segmento (grupo) específico. No entanto, dependendo do sector de actividade e da dimensão da empresa, podemos perguntar-nos se de facto esses nichos de mercado são interessante para ser explorados comercialmente.

Concluimos também que é possível criar segmentações com o mesmo número de segmentos mas que fornecem informação complementar. Por exemplo, a segmentação W10P forneceu resultados também interessantes e diferentes dos resultados obtidos com a segmentação K10P.

Relativamente às metodologias de selecção de segmentações, de um modo geral, conseguimos obter segmentações diferentes e muito interessantes de ponto de vista da análise com situações “atípicas” como é o caso da segmentação K10P em que cada cluster (segmento) se especializa claramente numa variável distinta de gastos ou ainda com a segmentação H5P pelos motivos referidos acima.

CAPÍTULO VII

“Success is a science; if you have the conditions, you get the result.”

(Oscar Wilde)

Resultados: Perspectiva de Negócio

Nesta parte do trabalho, vamos ilustrar como é possível analisar os resultados obtidos com o método de selecção de segmentações na perspectiva de negócio, sendo o domínio de aplicação escolhido o da área bancária. Em primeiro lugar, apresentaremos o contexto bancário e respectivos produtos (Secção 7.1). A seguir, descreveremos a oferta existente, no mercado nacional de cartões de crédito/débito, de duas entidades bancárias (Secção 7.2). Depois, analisaremos o alinhamento desses produtos com as segmentações obtidas (Secção 7.3). Ilustraremos, de seguida, como podem ser criados novos produtos, especificamente para segmentos que não são cobertos pela oferta existente (Secção 7.4). Finalmente, será apresentado um sumário da análise realizada (Secção 7.5).

7.1. Cartões Bancários

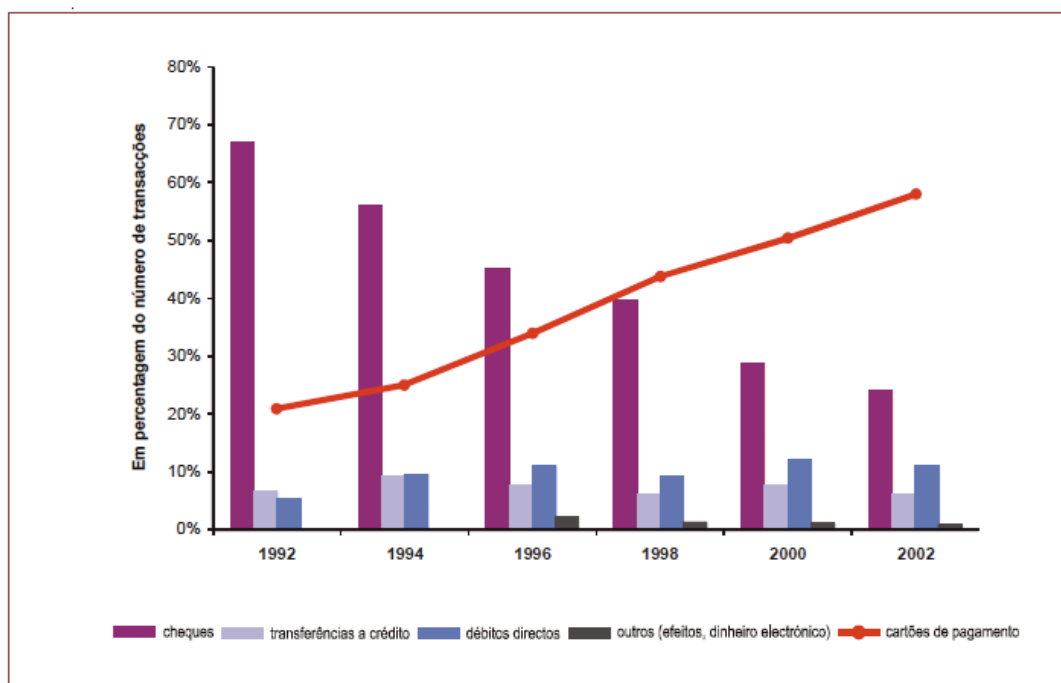
Os bancos disponibilizam uma grande variedade de produtos no mercado a fim de satisfazerem as necessidades dos consumidores. Os principais produtos bancários são o crédito hipotecário à habitação, o crédito ao consumo, e as contas de depósito. Os principais serviços são os cartões bancários, os cheques, as transferências, os créditos e os débitos directos (Banco de Portugal, 2009).

Os cartões bancários²⁹ são o instrumento de pagamento de bens e serviços mais utilizado em Portugal. Dos pagamentos que não utilizam numerário (notas e moedas),

²⁹ Os cartões bancários são instrumentos de pagamento emitidos por instituições de crédito ou sociedades financeiras devidamente autorizadas, que os disponibilizam aos titulares através da assinatura de um

mais de metade são, actualmente, efectuados com recurso aos cartões (*idem*, 2004)³⁰. A crescente utilização dos cartões bancários insere-se na tendência de evolução que se tem observado no nosso País nos últimos dez anos (ver figura 7.1.).

Figura 7.1. – Utilização dos instrumentos de pagamento em PORTUGAL (1992-2002)



Fonte: Caderno N° 6 do Banco de Portugal: “Cartões Bancários” (2004), disponível no site <http://www.bportugal.pt>

Abordaremos, de modo resumido, os diferentes tipos de cartões propostos pelos bancos. Existem, de um modo geral, três grandes tipos de cartões que são: o cartão de débito, o cartão de crédito e os cartões de pré-pagamento (*cash card*).

Originalmente, a diferença entre cartões de débito e de crédito era a de que o cartão de crédito permitia pagar mais tarde, enquanto que o cartão de débito correspondia a pagamento imediato. Actualmente, estes conceitos podem não estar tão claramente separados, já que existem cartões de crédito que permitem levantar dinheiro imediatamente e cartões de débito que dão dinheiro que ainda não tem disponível (*cash-advance*). Por exemplo, já existem cartões de débito (*e.g.* cartão M.T.V., descrito na

contrato (Decreto-lei n.º 166/95, de 15 de Julho e Aviso n.º 11/2001 do Banco de Portugal, de 6 de Novembro).

³⁰Fonte: Caderno N° 6 do Banco de Portugal: “Cartões Bancários” (2004) In. <http://www.bportugal.pt/pt-PT/PublicacoesIntervencoes/Banco/CadernosdoBanco/Biblioteca%20de%20Tumbnails/Cart%C3%B5es%20Banc%C3%A1rios.pdf>.

próxima secção) que possibilita pagamentos fraccionados, que disponibilizam uma linha de crédito paralela associada ao cartão, com prazo e taxa de juro próprios, disponível apenas para compras, para onde pode ser transferido o respectivo montante (normalmente valores elevados, por exemplo, viagens, computadores, L.C.D., etc.), libertando o *plafond* do cartão para as compras do dia-a-dia.

Em termos práticos, os cartões de débito podem ser utilizado para debitar pagamentos com os fundos disponíveis na conta. Por exemplo, levantar dinheiro, pagar serviços e carregar o telemóvel no Multibanco.

Um cartão de crédito permite antecipar a compra de um bem ou serviço sem ter de o pagar no imediato, o que pode ser muito útil e de grande ajuda nos meses mais difíceis. No entanto, compete à entidade emitente do cartão de realizar o pagamento das despesas efectuadas pelo titular do cartão, após o envio, pelo estabelecimento, das facturas relativas às aquisições realizadas. O pagamento realiza-se com a dedução de uma comissão correspondente ao serviço prestado pela entidade como contraprestação, por esta gerir o sistema e assumir os riscos. Apesar de ser uma criação recente, o cartão de crédito desponta como um dos mais utilizados meios de pagamento, além de instrumento de viabilização de crédito na sociedade actual. Adicionalmente, alguns cartões de crédito trazem outros benefícios aos clientes como, por exemplo: devolução de uma percentagem do que gastou em dinheiro, *milhas de viagens*, estadias em hotéis e descontos em lojas. Além disso, as instituições bancárias têm, normalmente, parcerias interessantes para os seus clientes, como seguros mais baratos. Obviamente, essas vantagens dependem de cada cartão, como veremos na secção seguinte.

Os cartões de pré-pagamento permitem fazer pagamentos (tipicamente de pequenas quantias) e levantamentos até ao limite do montante nele carregado, ou seja, são utilizados para pagamento de serviços específicos, relacionados, principalmente, com o uso de telefones e de meios de transporte públicos, ou compras de pequeno valor, por exemplo. Esse tipo de cartão pode ser recarregado várias vezes, observando-se, em cada uma delas, um valor limite de carregamento fixado pelo emissor.

Para aqueles três tipos de cartões existem, no entanto, algumas especificidades como se pode verificar na tabela 7.1..

Tabela 7.1. – Principais tipos de cartões e seu funcionamento³¹

	Tipo	Funcionamento
Cartões de pagamento (ou cartões de débito)	Cartão de autorização sistemática.	O saldo da conta é verificado em cada operação. É autorizado somente se o saldo da conta é suficiente e actualiza o saldo imediatamente.
Cartões de pagamento (ou cartões de débito)	Cartão de débito imediato.	Os pagamentos e as retiradas de dinheiro são possíveis até atingir o <i>plafond</i> de utilização. O débito em conta passa, geralmente, em 24 horas.
Cartões de pagamento (ou cartões de débito)	Cartão com débito diferido.	Os pagamentos e as retiradas de dinheiro são autorizados até atingir o <i>plafond</i> de utilização. Os débitos em conta são agrupados e afectados à conta uma vez por mês.
Cartão de crédito	Cartão associado a um crédito permanente (crédito <i>revolving</i> /crédito renovável)	O montante disponível sobre o crédito é geralmente verificado a cada operação. Se o saldo for suficiente, a operação é autorizada.
Cartão pré-pago	Pré-pagamento.	É um cartão que tem associado um montante pré-pago ou um saldo disponível no próprio cartão, normalmente limitado a determinado valor. Quando é utilizado, origina reduções no valor pré-pago ou no saldo disponível. Este tipo de cartões caracteriza-se por desempenhar funções pré-pagas (por exemplo, carregar o telemóvel). A filosofia deste cartão é carregá-lo com dinheiro para podê-lo usar depois. Ao contrário dos cartões de crédito, paga-se antes de usar.

Salientamos que os cartões têm ainda uma dimensão identificativa, porque apenas a sua posse legítima permite ao titular auferir dos benefícios inerentes ao sistema, provando ao comerciante a sua relação contratual com a entidade emitente.

A atribuição de cartões tem geralmente em conta o rendimento (elevado / baixo) e o nível de risco (elevado / baixo) do cliente (ver figura 7.2.), sendo que cada cliente tem necessidades diferentes.

³¹ Fonte: <http://www.lesclesdelabanque.com>.



Figura 7.2 – Avaliação dos clientes em função do rendimento e o nível de risco (Adaptado de Darnaoui, 2005)

É muito comum que haja uma adequação entre a oferta de cartão e algumas características simples do perfil da população-alvo, como se pode ver na tabela abaixo.

Tabela 7.2. – Atribuição de cartões em função de categorias de clientes³²

População Produto	Estudantes, jovens, reformados	Empregados e clientes com rendimentos fixos	Profissões liberais	Empresas	Clientes “VIP”
Cartão de autorização sistemática					
Cartão de débito imediato					
Cartão com débito diferido					
Cartão de crédito					
Cartão pré-pagamento					

No entanto, esta forma de atribuir os cartões aos clientes (ver tabela 7.2.), não tem em conta o mais importante que é o padrão de consumo de cada um. Para além do perfil socioeconómico e da faixa etária, é importante entender também a distribuição das despesas por diversas categorias. Por exemplo, um estudo de mercado poderia identificar perfis diferentes dentro de uma das categorias de clientes apresentadas na tabela 7.2. (por exemplo, estudantes que têm as despesas concentradas em comida e estudantes que têm as despesas concentradas em gasolina e entretenimento). Com base

³² Fonte: <http://www.lesclesdelabanque.com>.

nestas tendências, é possível desenvolver novos produtos, mais adequados para os segmentos respectivos. No presente trabalho, vamos mostrar como isso pode ser feito com base em perfis de consumo obtidos com segmentação dos clientes usando métodos de *clustering*.

Entendemos que para o alinhamento de cartões bancários, com os segmentos analisados anteriormente, seria interessante abordar duas perspectivas. Na primeira abordagem, pretendemos basear-nos nos cartões bancários existentes no mercado nacional e alinhá-los com as segmentações e respectivos segmentos que determinamos. Numa segunda abordagem, vamos supor que somos um banco e queremos desenvolver cartões de débito/crédito para os segmentos que não estejam cobertos por nenhum cartão existente e tentar desenvolver um produto específico para esses segmentos.

7.2. Cartões Bancários Seleccionados

A partir das segmentações geradas, vamos, nesta primeira abordagem, basear-nos na oferta de cartões bancários existentes no mercado nacional e adaptá-los aos segmentos obtidos. No entanto, para simplificar a análise, devido a vasta gama de produtos disponibilizados, analisaremos a oferta de cartões de duas instituições bancárias³³: o Santander Totta (S.T.) e a Caixa Geral de Depósitos (C.G.D.) com base nas segmentações obtidas no capítulo anterior.

No que diz respeito à oferta de cartões bancários propostos pelo Santander Totta (S.T.), foram seleccionados os cartões apresentados na tabela 7.3..

Tabela 7.3. – Produtos seleccionados do Santander Totta

Cartões	Tipo	Descrição	Anuidades e juros
	Crédito	É o cartão mais exclusivo do ST que permite o acesso a limites de crédito elevados, a um vantajoso pacote de seguros (saúde, acidentes pessoais e outros) e serviços de assistência. Oferece descontos no alojamento e 10% em restauração com as instituições aderentes (e.g. Pousadas de Portugal). Permite converter as suas compras em viagens (Programa milhas).	Anuidade: Para o 1º titular é de 80€ e 40€ para outros cartões. TAEG de 17,97%
	Crédito	O <i>Caixa Gold</i> é um cartão de prestígio, que lhe permite efectuar pagamentos a crédito, beneficiar de um completo pacote de seguros e obter descontos nos serviços associados (hotéis	Encargo anual de 62,40€ TAEG de 24,10%

³³ A informação foi obtida em 30/06/2009.

		e aluguer de automóveis). Acesso a parcerias exclusivas e nível de serviço superior.	
Premium Travel 	Crédito	O cartão <i>Premium Travel</i> permite-lhe converter as suas compras em viagens para qualquer destino à sua escolha (Programa milhas). Receba mais 50% de milhas, comparativamente aos outros cartões. O “cartão <i>Premium Travel</i> leva-o muito mais longe, porque cada 1€ equivale a 1,5 milhas” (S.T., 2009).	Taxa de emissão gratuita para o 1º e 2º Titular. O cartão a atribuir ao 1º titular tem uma anuidade de 65€/ano e ao 2º tem uma anuidade de 30€/ano, cobradas semestralmente. TAEG de 16,77%
Light 	Crédito	Cartão ideal para as suas necessidades e gostos, com a vantagem de lhe permitir escolher quando e como deseja pagar, com toda a flexibilidade. Tem a opção de pagamentos até 60 meses, com a taxa de juro mais baixa do mercado.	Anuidade de 16,5€, estando isentos se o valor de compras for maior ou igual a 1.200 euros, por ano. TAEG de 13,2% ³⁴ .
10.10 Tsi 	Crédito	Para fazer todas as suas compras e pagamentos e que lhe permite beneficiar de 10% de desconto em combustíveis e 10% de desconto em portagens. Para beneficiar dos descontos de 10% é necessário realizar compras a crédito com o mínimo de 200€/mês, não considerando os pagamentos de combustíveis e portagens.	Anuidade: A cobrança desta taxa é semestral com um valor de 13,75€. TAEG de 26,2%.
Cartão 5% 	Crédito	Um cartão inovador que lhe dá sempre 5% de desconto, sem limite de valor, em todas as suas compras. Mais concretamente, este cartão dá 5% de desconto sobre 50% do montante utilizado (em 1000€ utilizados tem 25€ de “bónus”) e sobre os restantes 50% paga 3% (juros), mas recebe 5%. Ou seja, paga 15€, mas recebe 25€ - resultado: recebeu um bónus líquido de 35€ sem quaisquer encargos.	A anuidade é grátis no 1º ano, estando isentos se o valor de compras for maior ou igual a 600€ por ano. Caso contrário, a anuidade será de 30€ para o 1º Titular e de 20€ para o 2º Titular. TAEG de 19,2%.
Electron 	Débito	Permite efectuar o acesso a várias operações, como levantamentos e transferências em caixas automáticos e pagamentos em terminais de pagamento automático (nas lojas). Cartão Multibanco, mas com a vantagem adicional de estarem associados à Rede Visa.	Este cartão é grátis ao efectuar 600€/ano, ou 50€/mês em movimentos.
Premier 	Débito	É um cartão de débito de prestígio que acresce ao cartão <i>Electron</i> o acesso à vários Seguros Associado (saúde, viagem, etc.). Oferece protecção às compras, assistência em viagem.	Encargo anual de 64,48€
Recarregáveis 	Pré-Pago	Cartão ideal para as compras ou despesas do dia-a-dia, para as férias ou ainda para oferecer como uma prenda em ocasiões especiais. Um cartão que pode dar aos seus filhos para pagarem a universidade, o colégio, os livros, o passe, etc. ou à empregada doméstica para pagar as compras do dia-a-dia para a sua casa	Sem montante mínimo de carregamento, excepto no 1º carregamento que terá de ser no valor de 25€.

Fonte: www.santandertotta.pt

³⁴ A Taxa Anual de Encargos Efectiva Global (TAEG). A TAEG é a taxa anual que expressa o custo total da utilização do seu cartão de crédito. Exemplo para uma utilização de crédito de 1.500€, com reembolso em 12 prestações mensais de 133,62€, incluindo impostos legais aplicáveis. O montante total imputado ao consumidor é de 1.603,40€ e o custo total do crédito de 103,40€.

Em relação aos cartões propostos pela Caixa Geral de Depósitos (C.G.D.)³⁵ foram seleccionados os seguintes cartões apresentados:

Tabela 7.4. – Produtos seleccionados da Caixa Geral de Depósitos

Cartões	Tipo	Descrição (resumida)	Anuidades e juros
	Débito diferido	Este cartão vai permitir-lhe, através da sua utilização, contribuir para a valorização das áreas de interesse da comunidade de fãs, nomeadamente no campo do desporto, da cultura e da responsabilidade social. Os pagamentos das compras são debitados mensalmente, de uma única vez, beneficiando da utilização de um <i>plafond</i> autorizado e sem quaisquer custos adicionais por este deferimento.	Encargo anual 9,50€.
	Débito	Permite movimentar a sua conta à ordem, em Portugal, no estrangeiro e na Internet. Pode também beneficiar de um pacote de serviços de assistência em viagem, totalmente gratuito.	Encargo anual de 8,50€, sendo gratuito no 1º ano.
	Débito diferido	Este cartão permite permitir o acesso ao Programa Farmácias Portuguesas (acumular pontos pelas compras de determinados serviços e produtos disponíveis na sua farmácia e trocá-los por serviços farmacêuticos e produtos de saúde e bem-estar). Permite usufruir de ofertas especiais, flexibilidade pagamento, diferimento no pagamento das compras nas farmácias aderentes.	Encargo anual 3€. Isenção da anuidade em função de volume médio mensal de compras e levantamentos a crédito de apenas 100 euros (1200 euros/ ano). Taxas de juro a partir de 12%.
	Débito Diferido	Permite efectuar compras com diferimento de pagamento e efectuar levantamentos directamente na tua conta à ordem. Ter mais de 15 anos. Vantagens: O acesso directo e gratuito a alguns bilhetes dos eventos da M.T.V., ter descontos nos bilhetes para festivais e outros espectáculos. Há ainda descontos em viagens, restaurantes, tecnologias, artigos de moda e desporto.	Encargo anual de 10€. Taxa de juro para versão com <i>revolving</i> ³⁶ é de 18,9%
	Débito	O <i>Megacartão Jovem</i> é um cartão para jovens dos 12 aos 25 anos, que dá descontos e funciona como cartão bancário. Dá acesso a todos os descontos que o Cartão Jovem oferece (viagens, espectáculos, compras e serviços). Oferece ainda descontos e seguros de acidentes pessoais e de assistência em	Encargo anual 8€.

³⁵ No entanto, devido a especificidades de alguns cartões da C.G.D. como, por exemplo, ter de ser sócio de uma determinada entidade ou ter um determinado estatuto profissional, salientando que alguns cartões são abrangidos pelos fenómeno de *cobranding*. Assim sendo, optámos por não incluir os cartões Benfica, LD<30, Maestro RE, LA Card Visa, Visabeira Exclusive, CUP, ITIC.

³⁶ Cartão com *revolving*, isto quer dizer que não tem de pagar na hora as compras. Aos maiores de 18, é dado um *plafond* previamente definido cujo pagamento deve ser feito a 100% no final de um período estabelecido.

		viagem gratuitos.	
Caixa Woman 	Débito / Crédito	Permite descontos até 5% sobre nas compras efectuadas em super e hipermercados. Pode utilizá-los para efectuar o pagamento de bens e serviços em lojas, restaurantes, supermercados, estações de serviços, entre outros num total de mais de 48.000 estabelecimentos comerciais em Portugal e de 23 milhões em todo o mundo.	Isenção das anuidades com uma utilização média mensal de apenas 200€, contando a utilização de ambos os cartões. Encargo anual de 20€. TAEG de 23,3%.
Caixa Carbono Zero 	Crédito	Permite de Beneficiar de condições especiais na aquisição de bens (e.g. L.C.D., computadores, etc.) e serviços mais eficientes no consumo de energia e com melhor desempenho ambiental. Uma solução que o ajuda a reduzir emissões e a poupar, enquanto poupa o ambiente.	Encargo anual 25€. Isenção da anuidade desde que apresente um volume anual mínimo de facturação de 3.000€. TAEG de 23,80%
Caixa Gold 	Crédito	Este cartão oferece condições vantajosas em várias áreas de interesse – viagens, cultura, estética e saúde, vestuário e joalharia, comunicações, assistência ao lar e seguros. E obter descontos nos serviços associados (hotéis e aluguer de automóveis). Acesso a parcerias exclusivas e nível de serviço superior.	Encargo anual de 65€. Oferta das anuidades com a simples utilização do cartão, num montante mensal médio igual ou superior a 750 euros. TAEG de 25,8%.
Caixa Classic 	Crédito	O cartão que lhe devolve até 3% do valor das suas compras, com um limite de 120€/trimestre. Permite uma linha de crédito gratuito até 45 dias após a realização das compras.	Encargo anual de 25€. Isenção das anuidades com uma utilização média igual ou superior a 250€. TAEG de 21,7%
Impar 	Crédito	Permite uma maior flexibilidade na gestão do orçamento familiar. Permite obter descontos em diversas unidades hospitalares e permite-lhe, ainda, beneficiar da oferta de um pacote de seguros e descontos em serviços associados (hotéis, aluguer de automóveis e outros). Crédito com taxa de juro reduzida sobre todo o saldo em dívida originado em compras.	Encargos anuais variam entre 10 e 40€ (em função da taxa de juro de crédito escolhida pelo cliente). TAEG a partir de 4,80% (para um capital de 2000€ a 12 meses, com 0% de juros, anuidade e impostos).
LEVE 	Crédito	É um cartão de poupança (P.P.R.) concebido a pensar naqueles que gostam de poupar para o futuro sem sacrificar o presente. O titular do cartão Leve pode usufruir de um aumento da sua poupança, com devolução até 1,5% do valor efectuado em compras para o PPR leve, com um máximo de <i>cash-back</i> mensal de 100€.	Encargo anual 18€. A isenção da anuidade tem por base o volume de facturação a crédito (compras e <i>cash-advance</i>), sendo o mínimo anual de 3.000€. TAEG de 23,80%
Miles & More 	Crédito	O cartão que oferece mais milhas e benefícios a quem viaja com frequência de avião. O cartão <i>Miles & More</i> da Caixa, nas versões <i>Gold</i> e <i>Classic</i> , emitido em parceria com a Lufthansa, permite acumular 5 milhas por cada múltiplo de 5€ em todas as compras ou levantamentos a crédito.	Encargo anual 30€ para a versão <i>Classic</i> e 70€ para a versão <i>GOLD</i> . TAEG de 24,12%
Soma 	Crédito	Permite converter as suas compras em pontos para o cartão <i>fast</i> da Galp Energia, beneficiar de um pacote de seguros e obter descontos nos serviços associados (Por ex. por cada 5€	Encargos anuais de 20€, isentos caso a facturação (compras e levantamentos a crédito)

		gastos com o cartão, dá 5 pontos no cartão <i>Fast Galp</i> e com o cartão <i>Fast Galp</i> , em troca de <i>X</i> pontos é possível ter bilhetes de cinema a 1€).	do ano anterior seja igual ou superior a 3000€. TAEG de 24,43%
LOL (Júnior / Lol / Pró) 	Pré-Pago	Um cartão “Livre, Original e Leve” (C.G.D., 2009). Cartão ideal para as compras na Internet. Em termos de funcionamento, é muito semelhante a um cartão de débito normal, mas pode usá-lo até o limite de fundos previamente carregados. No caso do Pró (para os mais de 26 anos), o montante do carregamento pode variar entre os 5€ e os 2.000€, enquanto no <i>Lol</i> (para os 15-25 anos) pode variar entre os 5€ e os 500€.	Encargos anuais de 7€, 3€ e 3€ respectivamente.

Fonte: www.CGD.pt

A escolha nos cartões está longe de ser simples. No entanto, tentámos, de forma simplificada, atribuir um possível estatuto ou “*ranking*” (em termos de gastos/despesas) dos seguintes cartões: cartão *Light*, cartões recarregáveis / *Lol* para o segmento baixo, cartão *Classic* (para o segmento médio), *GOLD* (para o segmento alto) ao *Platinum* do S.T. (para o segmento alto-luxo). Este *ranking* é importante para informar que a atribuição dos cartões também tem em conta o rendimento das pessoas.

De seguida, analisaremos cada cartão em termos das características que foram usadas nas segmentações no fim de perceber melhor o processo de alinhamento de produtos.

Em relação ao S.T., constata-se, como se pode ver na tabela 7.5., que não tem uma oferta de cartões específicos muito variada. Tendo em conta, em particular, os rendimentos (*e.g.* cartão *Light*, *Gold*, e *Platinum*) na estratégia da atribuição de cartões bancários.

Tabela 7.5. – Cartões do S.T. em termos das características usadas nas segmentações

Banco	Cartões	Características
S.T.	Platinum	O cartão <i>Platinum</i> é adequado para pessoas de alto poder aquisitivo, com perfil de gastos muito elevados e que viajam com frequência ao estrangeiro, necessitando utilizar serviços desses países, como seguros, médico, e automóveis.
	GOLD	Este cartão é direccionado para clientes com grande poder aquisitivo e que pretendem manter um relacionamento comercial privilegiado com a entidade bancária.
	Premium Travel	Este cartão é adequado para pessoas que viajam muito.
	Light	Este cartão é ideal para pessoas com um poder aquisitivo mais moderado / baixo e que se preocupam por exemplo com os juros, porque nem sempre conseguem pagar a dívida a 100%.

	10.10 Tsi	O Cartão 10.10 TSi tem um programa de Fidelização que atribui descontos em combustível e portagens Via Verde. Ou seja, este cartão é adequado para pessoas que usam / andam muito de carro.
	Cartão 5%	O Cartão 5% foi criado a pensar em todos os clientes que preferem pagar de forma fraccionada o saldo em dívida do seu cartão de crédito. Ideal para efectuar compras (<i>shopping</i>).
	Electron	Este cartão é ideal para compras do dia-a-dia de menor valor e para levantar dinheiro sem taxas.
	Premier	Este cartão é adequado para clientes que realizam viagens com uma frequência esporádica. E é ideal para as compras do dia-a-dia de menor valor.
	Recarregáveis	Cartão ideal para as compras ou despesas do dia-a-dia, sendo um óptimo cartão para os mais jovens.

No que diz respeito à C.G.D., verifica-se, de um modo geral (ver tabela 7.6.), que um dos pontos-chave da estratégia da C.G.D. é ter uma oferta de produtos e serviços bancários específicos.

Tabela 7.6. – Cartões da C.G.D. em termos de características usadas nas segmentações

Banco	Cartões	Caracterização
C.G.D.	Caixa Fã	Este cartão contribui para a valorização das áreas de interesse, nomeadamente no campo do desporto, da cultura e da responsabilidade social, ou seja, este cartão é adequado para as actividades de lazer.
	Caixautomática Electron / Maestro	Este cartão é ideal para compras do dia-a-dia de menor valor e para levantar dinheiro sem taxas adicionais.
	Farmácias Portuguesas	Este cartão é adequado para pessoas com muitas despesas de saúde; pode eventualmente ser adequado para pessoas que fazem despesas de beleza, dado que muitos dos produtos são vendidos em farmácias.
	MTV	O cartão MTV é adequado para jovens. O nome é do canal de televisão, permitindo usufruir vantagens ligadas à música (festivais, C.D., etc.).
	Megacartão Jovem	Este cartão é destinado para os jovens entre 15 e 25 anos.
	Caixa Woman	Como o nome indica, este cartão é especialmente dedicado para as mulheres activas. Ideal para efectuar compras (<i>shopping</i>).
	Caixa Carbono Zero	“O Cartão que enCaixa na Natureza”. Cartão ideal para pessoas que se preocupam com o ambiente.
	Caixa Gold	Clientes alvo: Clientes que procuram um meio de pagamento internacional de prestígio e que mantenham um relacionamento comercial privilegiado com a C.G.D..
	Caixa Classic	Este cartão é ideal para pessoas com um poder aquisitivo mais moderado. Ideal para efectuar compras (<i>shopping</i>).
	Impar	Este cartão é adequado para pessoas que necessitam frequentemente de recorrer ao crédito. Ideal para efectuar levantamentos de dinheiro a crédito (<i>cash-advance</i>).
	LEVE	Este cartão é adequado para pessoas que se preocupam com o seu plano de reforma P.P.R..
	Miles & More	Este cartão é adequado para pessoas que viajam muito.
	Soma	Este cartão é adequado para pessoas que usam / andam muito de carro.
	Lol (Júnior / Lol / Pró)	É um cartão que poderá servir necessidades muito específicas, especialmente nas viagens (transporte de valores), na Internet (segurança).

A seguir, para ilustrar o processo de alinhamento dos cartões bancários com as segmentações geradas, vamos analisar, como exemplos, as segmentações K10P, K5P e W10P.

7.3. Alinhamento de Cartões com os Segmentos Obtidos

Vamos começar por ilustrar o método de alinhamento com um exemplo simples. Seleccionámos uma segmentação em que cada segmento está essencialmente associado a uma variável.

7.3.1. Segmentação K10P

A segmentação K10P, caracteriza-se, como já foi mencionado anteriormente, por ser uma segmentação em que cada segmento se especializa claramente numa variável distinta de gastos, como se pode ver na tabela 7.7..

Tabela 7.7. – Identificação de possíveis perfis de consumo com a segmentação K10P

Segmentação K10P											
Cluster	Nr. Clientes	pCARS	pTRAVEL	pCLOTHES	pHOME	pBEAUTY	pFOOD	pKIDS	pEDUCULTURE	pHOBBIESFUN	pMONEY
1	2164	10%	2%	8%	8%	10%	50%	1%	2%	3%	5%
2	26	12%	3%	4%	8%	9%	24%	30%	2%	3%	5%
3	347	8%	2%	6%	7%	8%	19%	0%	1%	2%	47%
4	188	9%	43%	6%	7%	7%	18%	0%	2%	2%	4%
5	663	45%	2%	5%	8%	8%	21%	0%	1%	3%	6%
6	509	7%	2%	36%	9%	12%	23%	1%	2%	2%	5%
7	450	9%	2%	9%	8%	41%	22%	1%	2%	3%	5%
8	72	8%	3%	9%	9%	7%	21%	1%	35%	3%	4%
9	79	13%	2%	6%	6%	9%	17%	0%	2%	42%	5%
10	502	11%	3%	8%	36%	9%	21%	1%	2%	3%	6%

No que diz respeito à adequação dos cartões bancários, verificámos que para o primeiro segmento (S1), os cartões mais indicados são aqueles que permitem condições especiais em compras para a categoria comida. No caso do S.T., não existe uma oferta

específica para compras em comida. No entanto, o banco S.T. tem cartões mais genéricos como os cartões *Light* e o cartão 5% que podem adequar-se para este segmento. Tendo isto em conta, recomendámos o cartão *Light* aos clientes que usam pouco o cartão, enquanto que para os clientes que necessitam de utilizar um cartão frequentemente, o cartão 5% é uma óptima solução, pelo facto de poder tirar proveito do desconto de 5% nas compras efectuadas que esse cartão oferece. Ou seja, pensamos que o cartão 5% se enquadra melhor, para este perfil de clientes, do que o cartão *Light*, pelo facto de oferecer um desconto muito mais vantajoso para as compras do dia-a-dia como, por exemplo, nas compras em super e hipermercados (alimentação). No caso da C.G.D., e mais concretamente para os clientes (masculinos), o cartão ideal será o cartão *Classic*, enquanto que para as clientes, o cartão ideal será o cartão *Woman*, que oferecem vantagens como, por exemplo, descontos até 3% e 5% respectivamente em lojas (vestuário e acessórios) e em super e hipermercados (comida).

O segundo (micro-)segmento (S2) tem a particularidade de ter gastos muito significativos na categoria crianças. Verificámos também gastos significativos em comida. Salientamos que o S.T. e a C.G.D. não têm uma oferta de cartões específicos para esta categoria. Assim, pensamos que, no caso do S.T., seria aceitável propor um cartão mais genérico como o cartão 5% que devolve 5% nas compras efectuadas. Outra possibilidade seria de propor, em vez do cartão 5%, o cartão *GOLD*, caso o cliente esteja numa escala de gastos elevada. Em relação aos cartões da C.G.D., pensamos que o cartão de crédito *Woman* seria uma óptima escolha, pelo facto de oferecer, por exemplo, descontos até 5% exclusivos em lojas e em super e hipermercados.

No que diz respeito ao terceiro segmento (S3), constatámos gastos muito significativos na categoria dinheiro e significativos na categoria comida. O cartão ideal, no caso do S.T., seria o cartão *Light*, pelo facto de ser uma óptima escolha para efectuar levantamentos de dinheiro a crédito, por exemplo. Salientamos que é possível participar num sorteio para ganhar uma viagem por cada 500€ em compras ou levantamentos³⁷ efectuados com o cartão *Light*. No caso da C.G.D., seria interessante oferecer um cartão mais específico para levantar dinheiro, como o cartão *Impar*, sendo que permite efectuar levantamentos de dinheiro a crédito (*cash advance*) com taxas muito vantajosas, mas, na nossa opinião, o cartão *Caixa Classic* seria também outra possibilidade, permitindo

³⁷ Campanha de Verão apenas válida entre 15 de Julho e 15 de Setembro de 2009.

aos titulares usufruir de descontos até 3% nas compras realizadas em supermercados (alimentação). Salientamos que, para ambos os bancos, seria também interessante sugerir aos clientes um cartão de débito tipo *Electron*, por ser uma boa solução para as compras do dia-a-dia com menor valor.

Em relação ao quarto segmento (S4), verificamos gastos muito significativos na categoria viagem e significativos em comida, sendo que se verifica também uma escala de gastos significativa. Na nossa opinião, os cartões ideais seriam, em primeiro lugar, para o S.T., o cartão *Premium Travel*, e, em segundo lugar, no caso da C.G.D., o cartão *Miles & More*, porque ambos os cartões são cartões específicos para os clientes que gostam de viajar e permitem, por exemplo, converter compras efectuadas em pontos-milhas, no fim de obter descontos na compra de um bilhete de avião. Salientamos que seria também interessante de propor, para além destes cartões, um cartão *Electron*, pelo facto de ter também vantagens de isenção de taxas na compra de bilhetes *online* no caso da companhia *low cost Easyjet*³⁸, por exemplo.

No quinto segmento (S5), é possível observar gastos muito significativos em carros e gastos significativos em comida. Os cartões disponíveis para este tipo de perfil de gastos são o cartão *10.10 Tsi* do S.T. que oferece descontos de 10% em combustíveis e 10% de descontos em portagens, e o cartão *Soma* da C.G.D. que converte as compras em pontos para combustíveis nos postos de abastecimento e permite, por exemplo, obter descontos no aluguer de automóveis. Seria também interessante que ambos os bancos propusessem um outro cartão como o cartão de débito *Electron*, por exemplo, para satisfazer as necessidades mais diversas do dia-a-dia.

O sexto segmento (S6) distingue-se dos outros segmentos por ter gastos muito significativos na categoria roupas e gastos significativos nas categorias comida e beleza. Nenhum dos bancos analisados tem um cartão que se encaixe perfeitamente neste perfil. Em relação ao S.T., poderia propor o cartão *5%* que oferece descontos exclusivos de 5% nas compras realizadas em supermercados (comida), lojas (vestuário e acessórios) e perfumarias (produtos de beleza), por exemplo. No caso da C.G.D., pensamos que o banco poderia oferecer o cartão *Caixa Classic* para os clientes (do sexo masculino), sendo que para as clientes, o cartão de débito *Woman* seria ainda mais adequado do que o cartão *Caixa Classic* pelo facto de oferecer mais vantagens como, por exemplo,

³⁸ www.easyjet.com.

descontos em Spas, ginásios e ainda possibilitar um desconto até 5% sobre compras efectuadas em super e hipermercados. No caso das compras serem efectuadas a crédito, recomendamos o *pack* “*Woman*”, sendo que o facto de se propor este *pack*, se tornar desnecessário propor o cartão de débito *Electron* pelo facto do *pack* “*Woman*” já incluir um cartão de débito.

No que respeita ao sétimo segmento (S7), verificámos que este segmento se caracteriza por ter clientes de faixa etária compreendida entre os 40 e os 55 anos, e com uma escala de gastos baixa, sendo que se verificam, porém, gastos muito significativos na categoria beleza. Quer isto dizer que estamos perante um segmento com clientes com uma idade madura, com uma certa preocupação pelo seu bem-estar (*beautiful people*). No caso do S.T., não há um cartão particularmente adequado para este segmento, pelo que a melhor escolha seria o cartão 5%, porque permite aos clientes deste segmento usufruir de descontos exclusivos ao adquirirem produtos de beleza nas lojas, ao efectuarem o respectivo pagamento com o este cartão. Em relação ao banco da C.G.D., o cartão mais adequado seria o cartão *Woman* que tem a vantagem de devolver até 5% do valor nas compras realizadas como, por exemplo, em artigos de beleza. Este cartão permite também aos titulares dos cartões Caixa *Woman* beneficiar de 10% de desconto nos centros *BodyConcept*³⁹ - *que proporcionam experiências verdadeiramente únicas de bem-estar*. Salientamos também o facto de que o cartão *Farmácias Portuguesas* poderia ser também uma opção para este segmento, mas isso partindo do pressuposto, como mencionado anteriormente, de que os produtos relacionados com beleza, adquiridos nas farmácias como, por exemplo, produtos dermatológicos (cremes para a pele, etc.) estão abrangidos na campanha deste cartão.

Em relação aos oitavo e nono segmentos (S8 e S9), constatámos gastos significativos nas categorias cultura (S8), lazer (S9) e comida (ambos), sendo que a respectiva escala de gasto pode ser considerada moderada. O S.T. não tem uma oferta de cartões específicos para pessoas com despesas nas categorias cultura e lazer, assim sendo, a melhor solução seria a de recomendar o cartão 5%. No entanto, seria também possível propor o cartão *Light*, mas, na nossa opinião, não seria tão vantajoso como o cartão 5%. Em relação aos cartões da C.G.D., recomendámos o cartão *Fã* que, aderindo à comunidade *Fã de cultura*, permitirá ao cliente do oitavo segmento (S8) obter

³⁹ www.bodyconcept.pt.

informações específicas sobre eventos culturais (festivais, actuações de bandas, teatro, etc.) que se realizarão a nível nacional e beneficiarão de preços exclusivos. Para os clientes do nono segmento (S9), este cartão fornece vantagens exclusivas para actividades de lazer (nomeadamente no campo do desporto, da cultura e da responsabilidade social) oferecendo, por exemplo, 40% de desconto na aquisição de bilhetes para os espectáculos e exposições na bilheteira da Culturgest⁴⁰. Para os clientes mais jovens, recomendamos o *Megacartão Jovem* que oferece vantagens exclusivas, tais como descontos para eventos, cinema, museus e monumentos, e ainda descontos em livraria aderentes (e.g. Bulhosa).

Finalmente, para o décimo segmento (S10), constata-se gastos muito significativos na categoria lar e gastos significativos na categoria comida. Mais uma vez, nenhum dos bancos tem oferta específica para este grupo de clientes. Assim sendo, sugerimos, no caso do S.T., propor o cartão 5% que permite devolver 5% do valor das compras do dia-a-dia do lar, sendo que seria também de oferecer o cartão de débito exclusivo aos (potenciais) titulares do cartão 5% para as despesas de menor valor ou para simplesmente utilizá-lo para efectuar levantamentos de dinheiro, por exemplo. Em relação à C.G.D., a melhor solução seria de oferecer o cartão *Woman* pelo facto de ser um cartão ajustado ao perfil identificado, ao possibilitar descontos e benefícios em estética, desporto, assistência ao lar, arquitectura, decoração e jardins. O cartão *Woman* dispõe, por exemplo, neste caso, de ofertas exclusivas às clientes *Caixa Woman* como beneficiar de 50% de desconto nas lojas *5àSec*⁴¹, e, para serviços de arranjos e transformações em peças de roupa e do lar, as titulares dos cartões *Caixa Woman* beneficiam de 10% de desconto nas lojas *Retoucherie de Manuela*⁴², por exemplo. Para os clientes do sexo masculino, a melhor solução seria a de sugerir o cartão *Caixa Classic* permitindo a estes clientes aproveitar também descontos até 3% nas compras efectuadas (ideal para as despesas do lar).

De um modo geral, verificámos que existem segmentos para os quais existem cartões perfeitamente ajustados (e.g. S4, S5), outros, parcialmente cobertos pelos cartões (e.g. S2, S8 e S9 no caso do S.T.), como se pode ver na tabela 7.8..

⁴⁰ www.culturgest.pt.

⁴¹ www.5asec.pt.

⁴² www.retoucheriedemanuela.pt.

Tabela 7.8. – Alinhamento dos cartões do S.T. e da C.G.D. com a Segmentações K10P

Bancos	Cartões	K10P									
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S.T.	Platinum										
	GOLD										
	Premium Travel										
	Light										
	5%										
	10.10 Tsi										
	Electron										
	Premier										
	Recarregáveis										
C.G.D.	Caixa Fã										
	Caixautomática Electron / Maestro										
	Farmácias Portuguesas										
	MTV										
	Megacartão Jovem										
	Caixa Woman										
	Caixa Carbono Zero										
	Caixa Gold										
	Caixa Classic										
	Impar										
	LEVE										
	Miles & More										
	Soma										
	Lol (Júnior / Lol / Pro)										

Legendas: **amarelo** – possível **verde** – segmento coberto pelo cartão

XXX – possível (se pressuposto incluído na campanha do cartão)

A seguir, vamos analisar os segmentos mais interessantes das segmentações K5P e W10P que foram analisadas no capítulo anterior nas secções 6.4.3. e 6.4.7., respectivamente.

7.3.2. Segmentação K5P

No caso da segmentação K5P, apresentada na tabela 7.9., e discutida na secção 6.4.3., vamos ilustrar especificamente o primeiro segmento (S1). Nesta segmentação, verificámos, como mencionado no capítulo anterior, que todos os segmentos têm uma escala de gastos moderada (entre os 3,2 e os 3,4) e uma faixa etária compreendida entre os 40 e os 55 anos.

Tabela 7.9. – Identificação de possíveis perfis de consumo com a segmentação K5P

Segmentação K5P											
Cluster	Nr. Clientes	pCARS	pTRAVEL	pCLOTHES	pHOME	pBEAUTY	pFOOD	pKIDS	pEDUCULTURE	pHOBBIESFUN	pMONEY
1	1185	20%	6%	16%	7%	22%	19%	1%	3%	2%	4%
2	414	11%	2%	7%	6%	8%	19%	0%	1%	2%	43%
3	111	13%	3%	8%	6%	8%	18%	1%	3%	34%	5%
4	658	13%	5%	9%	33%	8%	20%	1%	3%	3%	5%
5	2632	13%	3%	9%	8%	10%	47%	1%	2%	3%	5%

No que diz respeito à adequação dos cartões ao primeiro segmento (S1), verificámos que este segmento tem gastos significativos nas categorias comida, beleza e carros, sendo que os gastos são muito homogéneos nas restantes categorias em relação à média. Como em vários dos segmentos já analisados, não há, em nenhum dos bancos analisados, um produto específico para este segmento. Assim, entendemos que os cartões mais adequados seriam o cartão 5% do S.T., pelo facto de devolver 5% nas compras realizadas (por exemplo, se o cliente efectuar uma compra de produtos de beleza de 500€, o S.T. devolve-lhe 25€), e o cartão *Caixa Classic* da C.G.D., pelo facto de devolver até 3% do valor das compras efectuadas. Apesar disto, para as clientes do sexo feminino da C.G.D., a melhor solução seria a de sugerir o cartão *Woman* que permite, por exemplo, ter descontos em Spas, ginásios e nas compras em supermercados. Salientamos que seria também possível propor um outro cartão, como, por exemplo, um cartão de débito *Caixautomática Electron* ou *Maestro* no caso da C.G.D.; mas, no caso do cartão 5% do S.T., o banco já oferece gratuitamente um cartão de débito exclusivo com anuidades gratuitas para utilizar nas compras de baixo valor do dia-a-dia.

Para os outros segmentos, recomendamos cartões genéricos dependendo do rendimento como o cartão 5% do S.T. e o cartão *Woman* da C.G.D., sendo que, no caso do S.T., se o rendimento for baixo, o cartão *Light*, senão o cartão 5%. Salientando uma excepção com o segundo segmento (S2), onde um cartão específico seria recomendado

como o cartão Ímpar da C.G.D. que permite, por exemplo, efectuar levantamentos de dinheiro a crédito com taxas vantajosas, como se pode ver na tabela 7.10..

Verifica-se, de um modo geral, que os cartões genéricos são perfeitamente ajustados a um ou mais segmentos, outros são apenas parcialmente ajustados (*e.g.* S2 e S3 no caso do S.T.).

Tabela 7.10. – Alinhamento dos cartões do S.T. e da C.G.D. com a Segmentação K5P

Bancos	Cartões	K5P				
		S1	S2	S3	S4	S5
S.T.	Platinum					
	GOLD					
	Premium Travel					
	Light					
	5%					
	10.10 Tsi					
	Electron					
	Premier					
	Recarregáveis					
C.G.D.	Caixa Fã					
	Caixautomática Electron / Maestro					
	Farmácias Portuguesas					
	MTV					
	Megacartão Jovem					
	Caixa Woman					
	Caixa Carbono Zero					
	Caixa Gold					
	Caixa Classic					
	Ímpar					
	LEVE					
	Miles & More					
	Soma					
	Lol (Júnior / Lol / Pro)					

Legendas: amarelo – possível verde – segmento coberto pelo cartão

7.3.3. Segmentação W10P

Na segmentação W10P, apresentada na tabela 7.11. e discutida na secção 6.4.7., analisámos os segmentos 2, 5, 8 e 9 por serem os segmentos mais interessantes do ponto de vista desta análise.

Tabela 7.11. – Identificação de possíveis perfis de consumo com a segmentação W10P

Segmentação W10P											
Cluster	Nr. Clientes	pCARS	pTRAVEL	pCLOTHES	pHOME	pBEAUTY	pFOOD	pKIDS	pEDUCULTURE	pHOBBIES FUN	pMONEY
1	147	7%	3%	5%	6%	56%	16%	0%	1%	2%	4%
2	476	7%	2%	5%	5%	4%	72%	0%	1%	1%	3%
3	431	13%	4%	7%	37%	7%	22%	1%	2%	3%	6%
4	251	8%	1%	5%	6%	5%	18%	0%	1%	2%	54%
5	154	8%	46%	8%	6%	8%	14%	0%	1%	2%	6%
6	566	48%	2%	5%	7%	7%	21%	0%	1%	2%	6%
7	259	6%	1%	43%	10%	11%	17%	1%	2%	3%	6%
8	39	10%	1%	4%	5%	6%	15%	0%	2%	52%	5%
9	1489	10%	4%	13%	10%	17%	30%	1%	4%	4%	7%
10	1188	14%	2%	8%	10%	11%	45%	1%	2%	3%	6%

Verificámos que o segundo segmento (S2) se caracteriza por ter, por um lado, gastos muito significativos na categoria comida e, por outro, gastos pouco significativos nas restantes categorias, tendo em conta a média. De acordo com essas características, nenhum dos cartões analisados é particularmente adequado para este segmento. Assim, o cartão que se adapta melhor, no caso do S.T., é o cartão 5% ao oferecer descontos exclusivos nas compras realizadas em supermercados. Quer isto dizer que é o cartão ideal para o *shopping*, mas, para além deste cartão, sugerimos que o banco ofereça também o cartão de débito exclusivo (para os titulares do cartão 5%) para efectuar as compras de menor valor, acrescentando que um cartão *Recarregável* não é de excluir, podendo, por exemplo, ser proposto aos clientes mais jovens ou no caso de se tratar de um casal com filhos, ser também possível propor um cartão *Recarregável* aos filhos⁴³. Em relação aos cartões da C.G.D., acreditamos que a melhor decisão seria a de oferecer o cartão *Classic* que permite obter, por exemplo, descontos nos supermercados e por se tratar de um óptimo cartão para efectuar levantamentos de dinheiro sem taxas adicionais. No caso de clientes do sexo feminino, seria preferível propor o cartão de débito *Woman* pelo facto de oferecer vantagens como, por exemplo, descontos até 5% em lojas e supermercados, sendo que, no caso de efectuar frequentemente compras a

⁴³ De acordo com Cota, (2005), a aposta nos jovens é a nova “tendência” dos bancos.

crédito, ser interessante oferecer o *pack “Woman”* que inclui o cartão de débito e o cartão de crédito mini complementar *Woman* sem custos de anuidade.

Relativamente ao quinto segmento (S5), observamos gastos muito significativos na categoria viagens e gastos relativamente significativos em comida e em roupas. Na nossa opinião, o Banco S.T. poderia propor, por exemplo, o cartão *Platinum* que permite converter as compras e pagamentos a crédito em viagens e permite ainda o acesso a um tratamento preferencial e a descontos exclusivos por exemplo, sendo que é também possível obter acesso a limites de crédito elevados e efectuar levantamentos de dinheiro a crédito, por exemplo. No entanto, uma óptima solução seria de sugerir o cartão *Premium Travel* devido às características mais exclusivas deste segmento. No caso da C.G.D., a melhor opção seria, sem dúvida, o cartão *Miles & More* por ser um cartão destinado a pessoas que gostam de viajar. Este cartão permite, por exemplo, converter compras efectuadas em pontos-milhas no fim de obter descontos na compra de um bilhete de avião. No entanto, seria também interessante propor, para além destes cartões, um cartão *Electron*, pelo facto de ter também vantagens de isenção de taxas na compra de bilhetes *online*, no caso da companhia *low cost Easyjet*⁴⁴

Ao analisar o oitavo (micro-)segmento (S8), podemos observar gastos muito significativos na categoria lazer, e também gastos significativos em comida, e ainda gastos relativamente significativos em carros. Neste caso, também, não há cartões adequados para este perfil, que é bastante específico. Em relação à adequação dos cartões do S.T., pensamos que a melhor decisão seria a de propor o cartão *GOLD* que permite também facilidade de pagamentos e levantamento de dinheiro, possibilitando, por exemplo, um crédito gratuito para um período máximo de 50 dias. No caso do C.G.D., seria uma boa solução de propor simplesmente o cartão *GOLD* que oferece descontos nas entradas em museus e espectáculos como, por exemplo, usufruir de descontos de 30% na compra de dois bilhetes para espectáculos realizados na Culturgest (lazer). Para os gastrónomos (*fin gourmet*), este cartão oferece também descontos exclusivos em lojas *gourmet* e nos restaurantes aderentes. Finalmente, proporciona ainda descontos em serviços associados como hotéis, aluguer de carros, entre outros. Salientamos que o cartão *Woman* seria também uma forte possibilidade, pelo facto de também oferecer descontos em supermercados e em estações de serviço, por exemplo.

⁴⁴ www.easyjet.com

Todavia, devido às características específicas desse segmento, poderíamos também propor, no caso de recusa do cartão *GOLD*, um *pack* de cartões como, por exemplo, o cartão *Caixa Classic* e o cartão de débito *caixautomática Electron*.

O nono segmento (S9) caracteriza-se por ser, em primeiro lugar, o maior segmento da segmentação W10P e, em segundo lugar, por ter uma escala de gastos significativa (3,6 em 5). No que diz respeito a distribuição dos gastos nas várias categorias, verificámos gastos muito significativos na categoria comida e gastos significativos nas categorias beleza e roupas, e gastos relativamente significativos na categoria lar. Tendo em conta estas especificidades, pensamos que a melhor opção seria, para ambos os bancos, a de propor o cartão *GOLD*, que oferece aos seus titulares descontos exclusivos como, por exemplo, 10% de desconto especiais nos tratamentos termais (por exemplo, nas Caldas da Rainha). O *Santander Gold*, por exemplo, tem um serviço 24 horas com envio a casa de canalizador, electricista, serralheiro, entre outros, sem cobrar a deslocação e fornece também uma Assistência Médico-Sanitária ao Lar. Salientamos que o cartão 5% seria também uma possibilidade para este segmento, mas tendo em conta a escala de gastos ser significativa, pensamos que o cartão *GOLD* seria uma melhor opção. Em relação à adequação com os cartões da C.G.D., na nossa opinião, a melhor opção seria a de propor o cartão *Caixa GOLD*, que oferece descontos em séries de tratamentos de corpo e rosto e em produtos de cosmética e dietética⁴⁵, sendo que possibilita também descontos exclusivo em clubes de *Fitness* da Solinca (Grupo Sonae)⁴⁶, por exemplo. Permite ainda ao titular deste cartão, usufruir de um conjunto de regalias e de descontos na aquisição de produtos e serviços em mais de 150 ópticas da rede *Institutoptico*⁴⁷. Seria também uma alternativa, para este segmento, o cartão *Woman*, para as clientes do banco da C.G.D.. Salientamos que, no caso de recusa eventual do cartão pelo cliente da C.G.D., seria também possível oferecer o cartão *Caixa Classic*, mas não seria tão vantajoso que o cartão *GOLD*, isto porque o cartão *GOLD* oferece, como vimos anteriormente, condições vantajosas em várias áreas de interesse – viagens, cultura, estética e saúde, vestuário e joalharia, comunicações, assistência ao lar e seguros, entre outros.

⁴⁵ www.clinicapersona.com.

⁴⁶ www.solinca.pt.

⁴⁷ www.institutoptico.pt.

De um modo geral, observámos que alguns cartões são perfeitamente ajustados a um ou mais segmentos, outros são apenas parcialmente ajustados e outros a nenhum. Por outro lado, alguns segmentos têm cartões que lhes são perfeitamente ajustados, outros parcialmente ou quase nada, como se pode ver na tabela 7.12..

Tabela 7.12. – Alinhamento dos cartões do S.T. e da C.G.D. com a Segmentação W10P

Bancos	Cartões	W10P									
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S.T.	Platinum										
	GOLD										
	Premium Travel										
	Light										
	5%										
	10.10 Tsi										
	Electron										
	Premier										
	Recarregáveis										
C.G.D.	Caixa Fã										
	Caixautomática Electron / Maestro										
	Farmácias Portuguesas										
	MTV										
	Megacartão Jovem										
	Caixa Woman										
	Caixa Carbono Zero										
	Caixa Gold										
	Caixa Classic										
	Impar										
	LEVE										
	Miles & More										
	Soma										
	Lol (Júnior / Lol / Pro)										

Legendas: **amarelo** – possível **verde** – segmento coberto pelo cartão

XXX – possível (se pressuposto incluído na campanha do cartão)

De seguida, vamos tentar criar um cabaz de ofertas ajustado a vários dos perfis encontrados, cujas necessidades não são totalmente cobertas pelos cartões existentes no mercado, abrindo, assim, uma nova oportunidade no desenvolvimento de novos produtos.

7.4. Desenvolvimento de Novos Cartões

A análise realizada anteriormente permite identificar alguns segmentos para os quais não existe uma oferta de cartões suficientemente específica por parte dos Bancos seleccionados para este estudo. Assim, existe uma oportunidade de criar novos produtos para esses segmentos de forma a melhor servir os clientes. Em seguida, vamos exemplificar produtos que poderiam ser criados especificamente para esses segmentos.

Na segmentação K5P, os clientes do primeiro segmento (S1) têm gastos distribuído em carros, vestuário, beleza e comida. Assim, no caso do banco S.T., proponho o cartão *AYA*⁴⁸ que permite ter descontos de 3% em todas as compras efectuadas, e 7% numa determinada área de interesse, ou seja, o titular pode escolher uma área de interesse entre várias propostas, tais como compras em supermercados e grandes superfícies, comunicações telefónicas, actividades de lazer, institutos de beleza, etc.. Este cartão não concorre com o cartão 5% dado que, embora tenha maiores descontos numa categoria específica de despesas, tem um desconto menor em todas as outras. Assim, o cartão *AYA* seria adequado para clientes com um perfil de compras em que uma categoria é dominante, enquanto que o cartão 5% poderia ser oferecido a clientes com um perfil de despesas mais homogéneo.

Os clientes do nono segmento (S9) da W10P, por exemplo, têm gastos muito elevados em comida e em beleza, vestuário e ainda gastos significativos em lar. Assim, proponho o cartão *HOME+* que oferece 1% de descontos em todas as compras efectuadas mensalmente, em supermercados e hipermercados, mais descontos exclusivos em artigos de vestuário e artigos de decoração para os membros da *HOME+*, no *site* do ClubeFashion⁴⁹, e ainda é enviado, mensalmente, a revista *HOME*, para divulgar e informar os membros *HOME+* dos descontos (lojas de vestuário, artigos de beleza, etc.) e das novas promoções exclusivas dos parceiros aderentes.

Na segmentação K10P, os clientes do décimo segmento (S10) têm gastos muito elevados nas categorias lar e comida, e ainda gastos relativamente significativos em carros em relação à média. Assim, sugiro o cartão *MODERN-LIVING* que oferece descontos exclusivos numa vasta gama de produtos para o lar, alimentares e para o automóvel, graças a parcerias realizadas com empresas de referência (e.g. Moviflor⁵⁰, AKI⁵¹, Continente e Galp), sendo que o serviço de entrega e montagem de mobiliário será gratuito para as compras pagas com o cartão *MODERN-LIVING*.

Salientamos que era ainda possível criar, por exemplo, o cartão *PETIT* a pensar na categoria crianças, com descontos exclusivos em lojas aderentes destinadas as crianças (e.g. Petit Patapon⁵²), acrescentando a possibilidade de descontos nos institutos

⁴⁸ Aya em árabe significa milagre.

⁴⁹ www.clubefashion.com.

⁵⁰ www.moviflor.pt.

⁵¹ www.aki.pt.

⁵² www.petitpatapon.pt.

de beleza aderentes, ideal para quem se preocupa com o seu bem-estar e com a sua família.

Os cartões descritos nesta secção pretendem apenas ilustrar como a informação obtida com base nas segmentações pode ser usada para identificar oportunidades para criação de novos produtos. No entanto, é preciso ter em conta que a criação de um novo cartão tem em conta muitos outros factores, por exemplo, o segundo segmento (S2) da K10P é um segmento muito específico de mercado de acordo com a dimensão do segmento. Por isso, pode não haver um mercado suficientemente grande para tornar o cartão *PETIT* num produto lucrativo.

7.5. Sumário

Neste capítulo, analisámos algumas das segmentações seleccionadas no capítulo anterior, numa perspectiva de negócio, em particular dos cartões bancários.

Em primeiro lugar, analisámos a adequação dos cartões de duas instituições bancárias a alguns dos segmentos gerados. De um modo geral, observámos que o Santander Totta tem uma oferta menos direccionada para segmentos específicos do que a Caixa Geral de Depósitos. Isso significa que, em muitos casos, o cartão mais adequado para os segmentos foi o cartão 5% do S.T..

Relativamente à C.G.D., constatámos, de um modo geral, que, graças à grande variedade de cartões disponíveis, foi possível encontrar cartões específicos para mais segmentos (sobretudo na segmentação K10P, ver anexo E). Ainda assim, o cartão *Woman* da C.G.D. foi o mais adequado para vários segmentos, sendo, neste último caso, segmentado também por sexo.

Verificámos também que os cartões *LEVE* e *Caixa Carbono Zero* da C.G.D. não estão adequados a nenhum dos segmentos de todas as segmentações geradas. Isto deve-se ao facto de as segmentações não serem feitas com base em variáveis relevantes para os clientes-alvo destes cartões.

Apesar dos dois bancos oferecerem uma grande variedade de cartões, constatámos, com base nos dados usados, que existem segmentos que não estejam (ou apenas parcialmente) cobertos por nenhum cartão existente.

Assim sendo, identificámos oportunidades de desenvolver um produto específico para esses segmentos. Propusemos, então, cartões bancários como, por exemplo, o cartão *HOME+* para clientes com despesas associadas à gestão de uma casa, e para desfrutar de vantagens contínuas é oferecido aos titulares do cartão uma revista mensal *HOME*, ou ainda o cartão *AYA* no qual o titular pode escolher uma área de interesse, entre várias propostas e assim usufruir de descontos específicos, respondendo, deste modo, melhor às necessidades específicas dos clientes.

CAPÍTULO VIII

*"An experience is never a failure because
it always serves to show something"*

(Thomas Alva Edison)

Conclusão e Trabalho Futuro

Este capítulo resume o trabalho apresentado nesta dissertação (Secção 8.1), destacando as principais contribuições alcançadas ao nível metodológico de análise de dados e tecendo considerações quanto às suas implicações práticas (Secção 8.2). A seguir, debatem-se algumas das limitações deste estudo (Secção 8.3). No final são feitas várias sugestões para trabalhos futuros (Secção 8.4).

8.1. Sumário

Na primeira parte da dissertação, foram apresentados os fundamentos teóricos da segmentação em *marketing* onde tentamos demonstrar a grande utilidade e aplicabilidade dos métodos e técnicas utilizadas em problemas de segmentação como o *Data Mining* (DM). Uma das aplicações mais comuns do DM em *marketing* é a segmentação de clientes. E para esse fim, são usados métodos de *clustering* (análise de clusters). Neste trabalho usámos métodos de análise de clusters para segmentar um conjunto de pessoas, de forma a identificar grupos homogéneos a nível psico-demográfico e de consumo.

Para a criação dos subgrupos (cluster) recorremos ao *SAS Enterprise Guide* que possibilita a utilização de vários métodos de *clustering* de uma forma simples, rápida e muito eficiente.

A abordagem seguida neste trabalho partiu do pressuposto que a criação de segmentações “diferentes” que agrupam os clientes de modo distinto é interessante quer

do ponto de vista académico, quer na perspectiva do negocio, porque representam novas perspectivas sobre um dado mercado. No entanto, dada a facilidade em criar segmentações, o processo para escolher/seleccionar as que são diferentes entre si é difícil e implica um grande investimento de tempo. Assim sendo, desenvolvemos uma metodologia (semi-) automática para eliminar segmentações redundantes, sendo que uma segmentação é considerada redundante se existir outra que agrupa os clientes de forma semelhante. Esta metodologia para a selecção de *clusterings* / segmentações, permitiu-nos comparar os *clusterings* e seleccionar as segmentações mais diferentes, assim, com mais potencial para serem interessantes.

A metodologia proposta foi aplicada aos dados indicados acima. Gerámos diferentes segmentações, obtidas usando conjuntos diferentes de variáveis, variando o método de clustering e/ou variando o número de clusters, dão perspectivas diferentes sobre o conjunto de clientes analisado. Em geral, observámos que as segmentações que foram depois seleccionadas com a metodologia proposta são diferentes entre si, sendo atingido, assim, o objectivo principal deste trabalho. Embora alguns segmentos gerados sejam pequenos, devemos notar que são, na sua maioria de tamanho aceitável, tendo em conta o tamanho da amostra. Adicionalmente, é preciso ter em conta que a amostra usada é relativamente pequena relativamente ao universo de clientes de bancos ou grandes empresas de distribuição. Assim, será de esperar que esses segmentos tenham, na realidade, uma dimensão bastante significativa.

Em seguida, procurámos avaliar as segmentações seleccionadas numa perspectiva de *marketing*, focando na oferta de cartões de duas instituições bancárias, o Santander Totta (S.T.) e a Caixa Geral de Depósitos (C.G.D.), uma privada e outra pública respectivamente.

Numa primeira fase analisámos a adequação desses cartões bancários aos segmentos obtidos. Constatámos, por exemplo, que o Santander Totta tem uma oferta menos direccionada para segmentos específicos do que a Caixa Geral de Depósitos. Enquanto que a C.G.D. tem uma oferta de produtos e serviços bancários muito específica e variada. No entanto, no caso da C.G.D., verificou-se também que alguns cartões não cobriam nenhum dos segmentos identificados, pelo facto de se tratarem de cartões muito específicos⁵³.

⁵³ De acordo com os dados utilizados neste trabalho.

De um modo geral, verificámos que os cartões genéricos (*e.g.* o cartão 5% do S.T. e o cartão *Woman* da C.G.D.) são os produtos que mais segmentos cobriram devido as suas características muito polivalentes, não sendo, no entanto, particularmente adequados para nenhum. Constatámos também, ao analisar a oferta de cartões de crédito em ambos os bancos, que no caso dos produtos dito genéricos existem um certa homogeneização na oferta desses produtos.

Numa segunda fase, procurámos desenvolver com características direccionadas para os segmentos que não estejam adequadamente cobertos por nenhum dos cartões existentes.

Assim sendo, propusemos cartões bancários como, por exemplo, o cartão *HOME+* para clientes (no caso do nono segmento da W10P por exemplo) com despesas associadas à gestão de uma casa e para desfrutar de vantagens contínuas é oferecido uma revista mensal *HOME*, ou ainda o cartão *AYA* no qual o titular pode escolher uma área de interesse entre várias propostas e assim usufruir de descontos específicos, respondendo assim melhor às suas necessidades específicas dos clientes (no caso do primeiro segmento da K5P por exemplo).

8.2. Contributo da Tese

Nesta dissertação, propusemos metodologias para a selecção de segmentações. Estas metodologias são uma ferramenta importante pela facilidade que há, hoje em dia, em obter muitas segmentações diferentes com recurso a ferramentas de *Data Mining*. Em seguida, demonstrámos a utilidade da técnica para o *marketing*, mostrando como pode ser usada para otimizar a oferta de cartões bancários.

Este trabalho contribuiu também para o aprofundamento da temática da segmentação com recurso a técnicas de DM, em particular no contexto do *marketing*.

Este trabalho contribui ainda para um maior entendimento da importância e do impacto da segmentação para a área bancária, na sua política comercial e de *marketing*, pois fornece uma revisão detalhada sobre o *marketing* no sector bancário e uma aplicação prática (*e.g.* alinhamento de produtos) de segmentação.

E finalmente, este trabalho tentou ainda dar um contributo para o aprofundamento de conhecimentos sobre a área do *Data Mining* para os profissionais de

gestão, sendo uma temática que cada vez mais se coloca perante os gestores de *marketing* / vendas devido às suas grandes vantagens na extracção de conhecimento nos dados.

8.3. Limitações do Estudo

Apesar dos resultados positivos obtidos, o estudo encontrou algumas dificuldades e limitações, por exemplo no que diz respeito às técnicas utilizadas para a segmentação. Cada técnica analítica apresenta simultaneamente potencialidades e limitações (Hair *et al.*, 1998). Mesmo com o uso difundido da segmentação como auxiliar a tomada de decisão, as técnicas de segmentação mais comumente utilizadas envolvem um número de decisões subjectivas, e os gestores e investigadores podem não apreciar plenamente a extensão no qual essas decisões afectam o resultado (Hoek *et al.*, 1996).

Esses autores salientam que se os gestores são inconscientes dessas limitações e que assumam que o estudo de segmentação melhorará as suas probabilidades de sucesso, os seus esforços podem tornar-se tão dispendiosos como a procura do Santo Graal, e “*just as futile*”. (Hoek *et al.*, 1996).

Relativamente à parte do trabalho sobre o alinhamento de produtos encontra-se limitado pela nossa falta de conhecimento relativamente:

- à(s) segmentação(ões) existente(s) no mercado (bancário);
- à natureza do negócio do Santander Totta e da Caixa Geral de Depósitos;
- à significância da amostra, relativamente ao universo de (potenciais) clientes da banca.

Convém também referir que o secretismo por parte das entidades bancárias, limitou a análise dos resultados obtidos como por exemplo a falta de validação dos resultados obtidos sobre o alinhamento de produtos por parte das entidades bancárias.

E finalmente, uma outra limitação muito importante e difícil de resolver está

relacionado com questões legais, no que diz respeito ao direito à privacidade e protecção dos dados dos clientes, que dificulta não só a obtenção de dados (reais) para a investigação académica, como a sua utilização para fins de *marketing*.

8.4. Trabalho Futuro

Acreditámos que este esforço é válido como contributo para uma melhor compreensão do caso em estudo. A riqueza da informação obtida através dos dados, e a limitação do presente estudo, permite continuar a investigação. Neste sentido podem ser feitas as seguintes propostas de investigação futura:

- a) Aperfeiçoar os métodos de selecção de segmentações/eliminação de segmentações redundantes propostos, com base nos resultados obtidos. Por exemplo, no caso da metodologia de selecção simples, seria interessante aperfeiçoar esta metodologia ao torná-la mais flexível e mais direccionada para as necessidades dos profissionais de *marketing*. Em relação à metodologia de selecção com o índices de similaridade, seria importante automatizar completamente todo o processo a fim de eliminar os eventuais erros humanos durante o processo de selecção.
- b) A resolução do problema da “instabilidade do segmento” (*e.g.* Wedel e Kamakura, 1999, Blocker e Flint, 2007), reflectido pelas mudanças nas necessidades dos clientes por exemplo. Uma solução interessante para um trabalho futuro seria de conseguir automatizar completamente todo o processo de selecção dos *clusterings* a fim de responder melhor as actualizações frequentes nas bases de dados.
- c) Integração sistemática com critérios de avaliação em *marketing*:
 - homogeneidade intra-segmento;
 - heterogeneidade inter-segmento;
 - segmentos mensuráveis e identificáveis;
 - segmentos acessíveis e operacionalizáveis;
 - segmentos suficientemente grandes para serem lucrativos;
 - valor potencial/retorno em termos de negócio.

A falta de ligação desses critérios é uma limitação importante. Podemos referir, por exemplo, o trabalho realizado por Rebelo (2006) para tentar resolver este problema.

- d) Abordar o método difuso (fuzzy): De acordo com Tam *et al.* (2003), este método justifica-se na segmentação em *marketing*, porque permite ao utilizador de encontrar um compromisso entre dois critérios, que são frequentemente contraditórios, a homogeneidade e tamanho dos segmentos. Isso, pelo facto de ter a capacidade de gerar uma matrizes de pertença de cada indivíduo de cada classe. Ou seja, este método permite que um indivíduo pertença à várias classes ao mesmo tempo segundo o seu grau de pertença. Para a aplicação prática deste método em trabalhos futuros sugerimos, por exemplo, o trabalho realizado pelos autores Tam *et al.* (2003) como um exemplo interessante de aplicação do método difuso na área bancária.
- e) Aprofundar as potencialidades das ferramentas do software *SAS* (por exemplo em relação ao *Enterprise Guide* ou ainda ao *Enterprise Miner*). Na nossa opinião seria interessante utilizar o código *SAS* (programação) para tentar automatizar a estratégia (algoritmo) desenvolvida no capítulo V relativamente a metodologia de selecção de segmentações usando os índices de similaridade por exemplo.

Referências Bibliográficas

AAKER, D. A, KUMAR, V. e DAY, G. S. (2001), “Marketing Research”, 7th ed., Wiley;

ACHROL, R. S., e ETZEL, M. J. (2003), “The structure of reseller goals and performance in marketing channels”, *Journal of the Academy of Marketing Science*, Vol. 31, Nº 2, pp. 146–162;

ACHROL, R. S., e KOTLER, P. (1999), “Marketing in the network economy”, *Journal of Marketing*, Vol. 63, Nº 4, pp. 146–163;

AGRAWAL, R., J. GEHRKE, J., GUNOPULOS, D. e RAGHAVAN, P. (1998), “Automatic subspace clustering of high dimensional data for data mining applications”, In. *Proceedings of the ACM International Conference on Management of Data*, Seattle, WA USA;

ÁGUAS, P. (2005), “Determinação dos segmentos de mercado prioritários: uma metodologia para destinos turísticos”, Tese de Doutoramento, Instituto Superior de Ciências do Trabalho e da Empresa (ISCTE), Lisboa;

ALDENDERFER, M. S. e BLASHFIELD, R. K. (1984), “Cluster analysis”, San Francisco: Sage Publications;

ALLAIRE, Y. (1972), “Some marketing thoughts on competition between banks and near banks”, *Institute of Canadian Bankers Review*, Vol. 5, Nº 1;

ALVES, R. e BELO, O. (2003), “Mineração de Dados em Sistemas Multidimensionais”, Relatório Técnico 001, Universidade do Minho, Braga;

ALPERT, N. (1972), “Personality and the determinants of adult choice”, *Journal of Marketing Research*, Vol. 9, pp. 89-92;

ARABIE, P. e HUBERT, L. (1994), "Cluster Analysis in Marketing Research", in R. P. Bagozzi (Eds.), *Advanced Methods of Marketing Research*, Blackwell, pp. 160-189;

ASSAEL, H. e ROSCOE, A. (1976), "Approaches to Market Segmentation Analysis", *Journal of Marketing*, Vol. 7 (October), pp. 67-75;

ATHANASSOPOULOS, A. (2000), "Customer satisfaction cues to support market segmentation and explain switching behavior", *Journal of Business Research*, Vol. 47, pp. 191-207;

BAKER, M. J. (2003), "The Marketing Book", 5th edition, Butterworth-Heinemann;

BALAKRISHNAN, P. V., COOPER, M. C., JACOB, V. S. e LEWIS, P. A. (1996), "Comparative performance of the FSCL neural net and K-means algorithm for market segmentation", *European Journal of Operational Research*, Vol. 93, N° 2, pp. 346-57;

BANNON, D. (2003), "Relationship Marketing and Politics", *Proceedings of the Political Marketing Conference*, Middlesex University, September 18th-20th;

BARNES, S. J., BAUER, H. H., NEUMANN, M. e HUBER, F. (2007), "Segmenting cyberspace: a customer typology for the Internet", *European Journal of Marketing*, Vol. 41, N° 1/2, pp. 71-93;

BARNETT, W. P., e MCKENDRICK, D. G. (2004), "Why are some organizations more competitive than others? Evidence from a changing global market", *Administrative Science Quarterly*, Vol. 49, N° 4, pp. 535-571;

BASU, S., DAVIDSON, I. e WAGSTAFF, K. L. (2009), "Constrained Clustering: Advances in Algorithms, Theory, and Applications", Chapman & Hall/CRC;

BEANE, T. P. e ENNIS, D. M. (1987), "Market segmentation: a review", *European Journal of Marketing*, Vol. 21, N° 5, pp. 20-42;

BERRY, M. e LINOFF, G. (2004), “Data Mining Techniques”, John Wiley;

BEZDEK, J. C. (1981), “Pattern Recognition with Fuzzy Objective Function Algorithms”, Plenum Press, New York;

BIEHL, M., HAMMER, B., VERLEYSSEN, M. e VILLMANN, T. (2007), “Similarity-based Clustering and its Application to Medicine and Biology”, *Summary and Abstract Collection, Dagstuhl Seminar Proceedings 07131*, ISSN 1862-4405;

BLOCKER, C. P. e FLINT, D. J. (2007), “Customer Segments as Moving Targets: Integrating Customer Value Dynamism into Segment Instability Logic”, *Industrial Marketing Management*, Vol. 36, Nº 6, pp. 810-822;

BOOMS, B., BITNER, M. (1981), “Marketing strategies and organization structures for service firms”, in Donnelly, J., George, W. (Eds), *Marketing of Services*, American Marketing Association, Chicago, IL;

BORDEN, N. H. (1964), “The Concept of the Marketing Mix”, *Journal of Advertising Research*, Vol. 4 (June), pp. 2-7;

BÖTTCHER, M., SPOTT M., NAUCK, D. e KRUSE, R. (2009), “Mining changing customer segments in dynamic markets”, *Expert Systems with Applications*, Vol. 36, Nº 1, pp. 155-164.

BOWEN, T. (1998), “Market segmentation in hospitality research: no longer a sequential process”, *International Journal of Contemporary Hospitality Management*, Vol. 10, Nº 7, pp. 289–296;

BLATTBERG, R. e SEN, S. K. (1976), “Market segments and stochastic brand choice models”, *Journal of Marketing Research*, Nº 13, pp. 34-45;

BRACHMAN, R. e ANAND T. (1996), “The Process of Knowledge Discovery in Databases”, *Advances in Knowledge Discovery and Data Mining*, pp. 37-57;

BRITO, C. M. (1998), “O Marketing da 3ª Vaga”, *Revista Portuguesa de Marketing*, Ano 2, Nº 6, pp. 75-80;

BROCHADO, A. (2002), “A Segmentação de Mercado: Bases de Segmentação e Métodos de Classificação. Aplicação ao Mercado de Vinho Verde”, Tese de Mestrado, Escola de Gestão do Porto, Universidade do Porto, Porto;

BRUNNING, E. R., KOVACIC, M. L. e OBERDICK, L. E. (1985), “Segmentation analysis of domestic airline passenger markets”, *Journal of the Academy of Marketing Science*, Vol. 13, Nº 1, pp. 17-31;

CALANTONE, R., e SAWYER, A.G. (1978), “The stability of benefit segment”, *Journal of Marketing Research*, Vol. 15, Nº 3, pp. 395-404;

CANT, M. C., STRYDOM J. W., JOOSTE C. J. e DU PLESSIS, P. J. (2009), “Marketing Management”, 5th Edition, Juta Academic;

CATHELAT, B. e WYSS, R. (1989), “A new segmentation tool for marketing in the 90s: the sociostyles system”, *ESOMAR, 42nd Conference*, Stockholm, pp. 207-20;

CATHELAT, B. (1994), “Socio-Lifestyles Marketing: The New Science of Identifying, Classifying and Targeting Consumers Worldwide”, Probus Professional Pub, Chicago, Illinois;

CESTRE, G., DARMON, R. (1998), “Assessing consumer preferences in the context of new product diffusion”, *International Journal of Research in Marketing*, Vol. 15, pp.123-35;

CHANARON, J.J., LUNG, Y. (1995), “L'économie de l'automobile”, La Découverte, Collection Repères, Paris;

CLÉMENT, V. e THONNAT, M. (1993), “A knowledge-base approach to integration of image processing procedures”, *CVGIP image understanding*, N° 27, pp. 166-184;

COBRA, M. (1992), “Administração de Marketing”, 2ª edição, São Paulo Atlas;

COLE, R. M. (1998), “Clustering with Genetic Algorithms”, M. Sc., Department of Computer Science, University of Western Australia, Australia;

CODRON, J. M., GIRAUD-HERAUD, E. e SOLER, L. G. (2003), “French Large Scale Retailers and New Supply Segmentation Strategies for Fresh Products”, ERS/USDA, Washington;

COLLICA, R. S. (2007), “CRM Segmentation and Clustering Using SAS Enterprise Miner”, SAS Publishing;

COTA, B.V. (2005), “A Emergência do Marketing Bancário: O Mercado Jovem e as Parcerias Estratégicas com Universidades”, Lisboa, Edições Universidade Lusíada;

CRATOCHVIL, A. (1999), “Data mining techniques in supporting decision making”, Tese de Mestrado, Universidade de Leiden, Leiden;

D'AVENI, R. A. (1995), *Hypercompetitive rivalries*, New York: Free Press;

DARNAOUI, B. (2005), “La gestion des demandes de cartes et des GAB”, *APTBEF*, Tunisie, in www.apbt.org.tn/fr/html/ecommerce/ecommerce.asp;

DAY, G.S. (1990), “Market Driven Strategy: Process for Creating Value”, The Free Press, New York, NY;

DE QUEIROZ, K., e GOOD, D. A. (1997), “Phenetic Clustering in biology: A critique”, *Quarterly Review of Biology*, Vol. 72, N° 1, pp. 3-30;

DEGROOT, M. H., (1986), "Probability and Statistics", Addison-Wesley Publishing Company, Pennsylvania;

DEMBY, E. (1974), "Psychographics and From Whence It Came" In. William Wells, ed., Life-Style and Psychographics, Chicago, *American Marketing Association*, pp. 9-30;

DENTON, L. e CHAN, K. A. (1991), "Bank selection criteria of multiple bank users in Hong Kong", *International Journal of Bank Marketing*, Vol. 9, Nº 5, pp. 23-34;

DESARBO, W. S. (1982), "GENNCLUS: New Models for General Nonhierarchical Clustering Analysis", *Psychometrika*, Vol. 47, Nº 4, pp. 449-475;

DESARBO, W. S. e MAHAJAN, V. (1984), "Constrained Classification: the Use of a Priori Information in Cluster Analysis", *Psychometrika*, Vol. 49, Nº 2, pp. 187-215;

DIAS, M. 2002, "Parâmetros na escolha de técnicas e ferramentas de mineração de dados", *Acta Scientiarum*, Maringá, Vol. 24, Nº 6, pp. 1715-1725;

DIBB, S. e SIMKIN, L. (1991), "Targeting, Segments and Positioning", *International Journal of Retail & Distribution Management*, Vol.19, Nº 3, pp. 4-16;

DIBB, S. e SIMKIN, L. (1997), "A program for implementing market segmentation", *Journal of Business & Industrial Marketing*, Vol. 12, Nº 1, pp. 51-65;

DIBB, S. (1998), "Market segmentation: strategies for success", *Marketing Intelligence & Planning*, Vol. 16, Nº 7, pp. 394-406;

DIBB, S. e WENSLEY, R. (2002), "Segmentation analysis for industrial markets: Problems of integrating customer requirement into operations strategy", *European Journal of Marketing*, Vol. 36, Nº 1/2, pp. 231-251;

DICKSON, P. R. (1982), "Person-situation: Segmentation's missing link", *Journal of Marketing*, Vol. 46, pp. 56-64;

DICKSON, P. R. e GINTER, J. I. (1987), "Market segmentation, product differentiation, and marketing strategy", *Journal of Marketing*, Vol. 51, pp. 1-10;

DOYLE, P. (2000), "Value Based Marketing", Wiley, Chichester;

DOYLE, P. e STERN, P. (2006), "Marketing Management and Strategy", 4th Edition, FT Prentice Hall;

DOUGLAS, S. P. (2001), "Exploring new worlds: The challenge of global marketing", *Journal of Marketing*, Vol. 65, N° 1, pp. 103-107;

DUDA, O. R., HART, E.P., e STORK, G. D. (2002), "Pattern Classification", John Wiley and Sons, New York;

DUNN, J. C. (1974), "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", *Journal of Cybernetics*, Vol. 3, N° 3, pp. 32-57;

EDRIS, T. e ALMAHMEED, M. (1997), "Services considered important to business customers and determinants of bank selection in Kuwait: a segmentation analysis", *International Journal of Bank Marketing*, Vol. 15, N° 4, pp.126-133;

EIGLER, P. e LANGEARD, E. (1987), "Servuction: Le Marketing des Services", McGraw-Hill, Paris;

EISENHARDT, K. M., e MARTIN, J. A. (2000), "Dynamic capabilities: What are they?", *Strategic Management Journal*, Vol. 21, N° 10/11, pp. 1105-1121;

EVANS, F. B. (1959), "Psychological and objective factors in the prediction of brand choice", *Journal of Business*, Vol. 33, Nº 4, October, pp. 340-369;

EVERITT, B. S. (1993), "Cluster Analysis", Edward Arnold;

FASULO, D. (1999), "An analysis of recent work on clustering algorithms", Technical Report, Dept. Computer Science and Engineering, Univ. Washington;

FAYYAD, U., PIATETSKY-SHAPIRO, G. e SMITH, P. (1996), "The KDD process for extracting useful knowledge from volumes of data", *Communications of the ACM*, Vol. 39, Nº 11, pp. 27-34;

FERREIRA, C., FERNANDES, H., ALVES, V. e SANTOS, M. (2006), "O *Data Mining* na Compreensão do Fenómeno da Dor: Uma Proposta de Aplicação", Universidade do Minho, in *Actas da 1ª Conferência Ibérica de Sistemas e Tecnologias de Informação*, Esposende;

FORSYTH, D. e PONCE, J. (2003), "Computer Vision: A Modern Approach". New Jersey: Prentice-Hall;

FLODHAMMAR, A. (1988), "A salesforce approach to industrial marketing segmentation", *Quarterly Review of Marketing*, Nº 13, pp. 5-9;

FRANK, R. e STRAIN, A. (1972), "A segmentation research design using consumer panel data", *Journal of Marketing Research*, Vol. 9, pp. 385-90;

FRANK, R., MASSY, W. e WIND, Y. (1972), "Market Segmentation", Prentice-Hall, Inc., Englewood Cliffs, NJ;

FREATHY, P. e O'CONNELL, F. (2000), "Strategic reactions to the abolition of duty free: Examples from the European airport sector", *European Management Journal*, Vol. 18, Nº 6, pp. 638-645;

FREYTAG, P. V. e CLARKE, A. H. (2001), “*Business to Business Market Segmentation*”, *Industrial Marketing Management*, Vol. 30, N° 6, pp. 473-486;

GEHRT, K., SHIM, S., (2003), “Situational segmentation in the international marketplace: the Japanese snack market”, *International Marketing Review*, Vol. 20, N° 2, pp. 180-194;

GINSBURGH, V. e WEBER, S. (2002), “Product lines and price discrimination in the European car market”, *The Manchester School*, Vol. 70, N° 1, pp. 101-114;

GIUDICI, P. (2003), “Applied Data Mining: Statistical Methods for Business and Industry”, John Wiley & Sons, Ltd;

GOI, C. L. (2009), “A Review of Marketing Mix: 4Ps or More?”, *International Journal of Marketing Studies*, Vol. 1, N° 1, pp. 1-15;

GOLLER, S., HOGG, A., e KALAFATIS, S. P. (2002), “A new research agenda for business segmentation”, *European Journal of Marketing*, Vol. 36, N° 1/2, pp. 252–271

GREEN, P. E. (1977), “A new approach to market segmentation”, *Business Horizons*, February, pp. 61-6;

GREEN, P. E., TULL, D. S. e ALBAUM, G. (1988), “Research for Marketing Decisions”, 5th ed., Prentice-Hall, London, Chap. 17;

GROTH, R. (2000), “Data Mining: Building Competitive Advantage”, Prentice Hall PTR, USA;

GROVER, R. e SRINIVASAN, V. (1989), “An approach for tracking within-segment shifts in market shares”, *Journal of Marketing Research*, Vol. 26, N° 2, pp. 230-236;

HAIR, J., ANDERSON, R., TATHAM, R. e BLACK, W. (1998), “Multivariate Data Analysis”, 5th edition, Prentice-Hall, Englewood Cliffs, New Jersey;

HALEY, R. I. (1968), “Benefit segmentation: A decision-oriented research tool”, *Journal of Marketing*, Vol. 32, Nº 3, pp. 30-35;

HAN, E-H., KARYPIS, G., KUMAR, V., e MOBASHER, B. (1997), “Clustering Based on Association Rule Hypergraphs”, in *Proceedings of the ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, Montreal, Canada, June, pp. 9-13;

HAN, J. e KAMBER, M. (2006), “Data Mining: Concepts and Techniques”, 2nd edition, Morgan Kaufmann Publishers;

HAND, D. J., MANNILA, H. e SMYTH, P. (2001), “Principles of Data Mining”, Cambridge, MA: MIT Press;

HANLAN, J., FULLER, D. e WILDE, S. (2006), “Segmenting Tourism Markets: a critical review”, in P. Tremblay & A Boyle (eds), *Proceedings of To the city and beyond: Council for Australian University Tourism and Hospitality Education (CAUTHE) Conference*, Melbourne, Vic., Victoria University, Melbourne, Vic;

HARRISON, T. H. (1998), “Intranet Data Warehouse”, São Paulo: Editora Berkeley Brasil;

HARTIGAN, J. A. e WONG, M. A. (1979), “A K-means clustering algorithm”, *Applied Statistics*, Vol. 28, pp. 100–108;

HARVARD BUSINESS ESSENTIALS (2006), “Marketer's Toolkit: The 10 Strategies You Need to Succeed”, *Harvard Business School Press*, Boston, Massachusetts, chap. 4;

HOEK, J., GENDALL, P. e ESSLEMONT, D. (1996), “Market segmentation: A search for the Holy Grail?”, *Journal of Marketing Practice: Applied Marketing Science*, Vol. 2, Nº 1, pp. 25-34;

HOOLEY, G., SAUNDERS, J. e PIERCY, N. F. (1998), “Marketing Strategy and Competitive Positioning”, 2nd edition, Prentice Hall,;

HRUSCHKA, E. R. (1986), “Market Definition and Segmentation Using Fuzzy Clustering Methods”, *International Journal of Research in Marketing*, Vol. 3, Nº 2, pp 117-134;

HRUSCHKA, E. R., e EBECKEN, N. F. F. (2003), “A Genetic algorithm for cluster analysis”, *Intelligent Data Analysis*, Vol. 7, Nº 1, pp. 15-25;

JAIN, A. K., PINSON, C., e MALHOTRA, N. K. (1987), "Customer loyalty as a construct in the marketing of banking services", *International Journal of Bank Marketing*, Vol. 5, Nº 3, pp.49-72;

JAIN, A. K. e DUBES, R. C. (1988), “Algorithms for Clustering Data”, Prentice Hall;

JAIN, A. K., MURTY, M. N. e FLYNN, P. J. (1999), “Data Clustering: A Review”, *ACM Computing Surveys*, Vol. 31, Nº 3, pp. 264-323;

JOHNSON, S. C. (1967), "Hierarchical Clustering Schemes", *Psychometrika*, Vol. 2, pp. 241-254;

JOSHI, A. W., e CAMPBELL, A. J. (2003), “Effect of environmental dynamism on relational governance in manufacturer–supplier relationships: A contingency framework and an empirical test”, *Journal of the Academy of Marketing Science*, Vol. 31, Nº 2, pp. 176–188;

JUSTINO, L. (2007), “Direcção Comercial”, Lidel Edições técnicas;

KALWANI, M. U. e MORRISON, D. G. (1977), "A parsimonious description of the Hendry system," *Management Science*, Vol. 23, N° 5, pp. 466-77;

KALWANI, M. U. e SILK, A. J. (1982), "Structure of repeat buying for new packaged goods", *Journal of Marketing Research*, Vol. 17, N° 3, pp. 371-385;

KANTARDZIC, M. (2003), "Data Mining: Concepts, Models, Methods, and Algorithms", Wiley-IEEE Press;

KAUFMAN, L. e ROUSSEEUW, P. J. (1990), "Finding Groups in Data: An Introduction to Cluster Analysis", J. Wiley and Sons, Inc., New York;

KAYNAK, E., KUCUKEMIROGLU, O. e ODABASI, Y. (1991), "Commercial bank selection in Turkey", *International Journal of Bank Marketing*, Vol. 9, N° 4, pp. 30-9;

KAYNAK, E. e HARCAR, T. (2005), "American consumers' attitudes towards commercial banks: A comparison of local and national bank customers by use of geodemographic segmentation", *International Journal of Bank Marketing*, Vol. 23 N° 1, pp. 73-89;

KELLY, I. e NANKERVIS, T. (2001), "Visitor Destinations", Wiley & Sons, Milton;

KEEGAN, W. J. e GREEN, M. C. (1999), "Principles of Global Marketing", Prentice-Hall;

KERSTETTER, D., CONFER, J., e BRICKER, K. (1998), "Industrial heritage attractions: Types and tourists", *Journal of Travel and Tourism Marketing*, Vol. 7, N° 2, pp. 91-104;

KIM, S., JUNG, T., SUH, E. e HWANG, H. (2005), "Customer segmentation and strategy development based on customer", *Expert Systems with Applications*, Vol. 31, pp.101-107;

KONONENKO, I. e KUKAR, M. (2007), “Machine Learning and Data Mining: Introduction to Principles and Algorithms”, Horwood Publishing;

KOTLER, P. e DUBOIS, B. (2000), “Marketing Management”, 10th edition, Publi-Union Éditions, Paris;

KOTLER, P., WONG, V., SAUNDERS, J. e ARMSTRONG, G. (2005), “Principles of Marketing”, 4th European Edition, Prentice Hall, London;

KOTLER, P. (2003), “Marketing Management”, 11th Edition, Prentice Hall International Editions;

KREMERS, S. P. J., DE BRUIJN, G. J., SCHAAALMA, H. J. e BRUG, J. (2004), “Clustering of energy balance related behaviours and their intrapersonal determinants”, *Psychology and Health*, Vol. 19, N° 5, pp. 595-606;

KRISHNAMURTHI, L. e RAJ, S. P. (1991), “An empirical analysis on the relationship between brand loyalty and consumer price elasticity”, *Marketing Science*, Linthicum, Vol. 10, N° 2, pp. 172-183;

LAROCHE, M. e MANNING, T. (1984), "Consumer bank selection and categorization processes: a study of bank choice", *International Journal of Bank Marketing*, Vol. 2, N° 3, pp. 3 – 21;

LAROCHE, M., ROSENBLATT, J. A. e MANNING, T. (1986), “Services used and factors considered important in selecting a bank: an investigation across diverse demographic segments”, *International Journal of Bank Marketing*, Vol. 4, N° 1, pp. 35-55;

LAW., M. H. C., FIGUEIREDO, M. A. T. e JAIN, A. K. (2004), “Simultaneous Feature Selection and Clustering Using Mixture Models”, *IEEE Transactions on Patterns Analysis and Machine*, Vol. 26, N° 9, pp. 1154-1166;

- LAZER, W. (1963), "Lifestyle concepts and marketing", in Greyser *et al.* (Eds), *Toward Scientific Marketing*, AMA, Chicago, IL;
- LEGOHEREL, P. e WONG, K. K. F. (2006), "Market segmentation in the tourism industry and consumers spending: What about direct expenditures?", *Journal of Travel & Tourism Marketing*, Vol. 20, pp.15-30;
- LEISEN, B. (2001), "Image segmentation: the case of a tourism destination", *The Journal of Services Marketing*, Vol. 15, pp. 49-56;
- LERER, L. (2002), "Pharmaceutical marketing segmentation in the age of the Internet", *International Journal of Medical Marketing*, Vol. 2, N° 2, pp. 159-166;
- LILIEN, G. e KOTLER, P. (1983), "Marketing Decision Making – A Model-Building Approach", Harper & Row, New York, NY;
- LILLESAND, T. M., KIEFER, R. W., e CHIPAN, J. W. (2004), "Remote sensing and interpretation", 5ed. Madison: Wiley;
- LINDGREEN, A. e ANTIOCO, M. (2005), "Customer relationship management: the case of a European bank", *Marketing Intelligence & Planning*, Vol. 23, N° 2, pp. 136-154;
- LIU, H. e ONG, C. (2006), "Variable selection in clustering for marketing segmentation using genetic algorithms", *Expert Systems with Applications*, Vol. 34, pp. 502–510;
- LOVELOCK, C. (1996), "Services Marketing", 3rd edition, Prentice Hall, Englewood Cliffs, NJ;
- MAHAJAN, V. e JAIN, A. K. (1978), "An Approach to Normative Segmentation", *Journal of Marketing Research*, Vol. 15, N° 3, pp. 338-345;

MATSUYAMA, T. (1989), "Expert Systems for Image Processing: Knowledge-Based Composition of Image Analysis Processes", *CVGIP*, Vol. 48, N° 1, pp. 22-49;

MCDONALD, M. e DUNBAR, I. (2004), "Market Segmentation: How to do it, How to profit from it", Elsevier Butterworth-Heinemann publications;

MACQUEEN, J. (1967), "Some methods for classification and analysis of multivariate observations", in *Proceedings of Fifth Berkeley Symposium on Math. Statist. and Prob.*, pp. 281-297;

MÄENPÄÄ, K. (2006), "Clustering the consumers on the basis of their perceptions of the Internet banking services", *Internet Research*, Vol. 16, N° 3, pp. 304-322;

MAHAJAN, V. e JAIN, A. K. (1978), "An approach to normative segmentation", *Journal of Marketing Research*, Vol. 15, pp. 338-45;

MANNILA, H. (1997), "Methods and problems in data mining", *Proceedings of the 6th International Conference on Database Theory*, pp. 41-55;

MATEAR, S. e GRAY, R. (1995), "Benefit-based segments in a freight transport market", *European Journal of Marketing*, Vol. 29, N° 12, pp.43-58;

MATSUYAMA, T. (1989), "Expert systems for Image Processing: Knowledge-based Composition of Image Analysis Processes", *CVGIP*, Vol. 48, N° 1, pp. 22-49;

MENTZER, J. T., MYERS, M. B. e CHEUNG, M-S. (2004), "Global market segmentation for logistics services", *Industrial Marketing Management*, Vol. 33, N° 1, pp. 15-20;

MICHALSKI, R. S. e STEPP, R. (1983), "Learning from observation: Conceptual clustering", in MICHALSKI, R. S., CARBONELL, J. G., e MITCHELL, T. M. (Eds.), "Machine learning: An artificial intelligence approach", San Mateo, CA: Morgan

Kaufmann;

MO, C-M., HAVITZ, M. E. e HOWARD, D. R. (1994), “Segmenting Travel Markets with the International *Tourism* Role (ITR) Scale”, *Journal of Travel Research*, Vol. 33, Nº 1, pp. 24-30;

MOBASHER, B., COOLEY, R. e SRIVASTAVA, J. (2000), “Automatic personalization based on Web usage mining”, *Communications of the ACM*, Vol. 43, Nº 8, pp. 142-151;

MORIARTY, R. T. e REIBSTEIN, D. J. (1986), “Benefit segmentation in industrial marketing”, *Journal of Business Research*, Vol. 14, Nº 6, pp. 463-486;

MYERS, J. (1976), “Benefit structure analysis: a new tool for product planning”, *Journal of Marketing*, Vol. 40, pp. 23-32;

NEIDELL, L. A. (1970), “Procedures and Pitfalls in Cluster Analysis”, *Proceeding*, Fall Conference, American Marketing Association, Vol. 107;

OEHLER, K. L. e GRAY, R. M. (1995), “Combining Image Compression and Classification Using Vector Quantization”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, Nº 5, pp. 461-473;

PAÇO, A. (2002), “A Segmentação do Mercado Consumidor: Uma Aplicação ao Caso do Vestuário”, Tese de Mestrado, Universidade da Beira Interior, Covilhã;

PAS, E. I. e Huber, J. C. (1992), “Market Segmentation Analysis of Potential Inter-city Rail Travelers”, *Transportation*, Vol. 19, pp. 177-196;

PIATETSKY-SHAPIO, G. (1991), “Discovery, Analysis, and Presentation of Strong Rules”, *Collection of Knowledge Discovery in Databases*, pp. 229-248;

- PIERCY, N. F. e MORGAN, N. A. (1995), "Market segmentation: strategic and operational questions in implementation", *Marketing Education Group Conference Proceedings*, pp. 636-645;
- PLUMMER, J. T. (1974), "The concept and application of life style segmentation", *Journal of Marketing*, Vol. 38, N° 1, pp. 33-37;
- PRIDE, M. V. e FERRELL, C. O. (1987), "Marketing: Basic Concepts and Decisions", Houghton Mifflin, Boston, MA;
- PRIETO M. (2005), "Segmentation du Marché Automobile Européen : une Analyse Empirique", *XVèmes Journées du Séminaire d'Etudes et de Statistiques Appliquées à la Modélisation en Économie (SESAME)*, Septembre, Rennes;
- PUNJ, G. e STEWART, D. W. (1983), "Cluster Analysis in Marketing Research: Review and Suggestions for Application", *Journal of Marketing Research*, Vol. 20, N° 2, pp. 134-148;
- QUINN, L., HINES, T. e BENNISON, D. (2007), "Making sense of market segmentation: a fashion retailing case", *European Journal of Marketing*, Vol. 41, N° 5-6, pp. 439-465;
- RAAIJ, F. W. e VERHALLEN, M. T. (1994), "Domain specific market segmentation", *European Journal of Marketing*, Vol. 28, N° 10, pp. 49-66;
- RAND, W. M. (1971), "Objective criteria for the evaluation of clustering methods", *Journal of the American Statistical Association*, Vol. 66, pp. 846-850;
- RAO, V. e WINTER, F. (1978), "An application of the multivariate Probit model to market segmentation and product design", *Journal of Marketing Research*, Vol. 15, pp. 361-368;

- RAO, P. C. e WANG, Z. (1995), “Evaluating alternative segmentation strategies in standard industrial markets”, *European Journal of Marketing*, Vol. 29, Nº 2, pp. 58-75;
- REBELO, M. (2006), “Segmentação do Comportamento Online utilizando Clickstream Data”, Tese de Mestrado, Universidade do Porto, Porto;
- REIS, E. (2001), “Estatística Multivariada Aplicada”, 2ª edição, Edições Sílabo, Lisboa;
- RICHERS, R. e LIMA, C. P. (1991), “Segmentação”, São Paulo: Nobel;
- ROBINSON, J. A. (1966), “Category clustering in free recall”, *Journal of Psychology*, Vol. 62, Nº 2, pp. 279-285;
- RODRIGUEZ, S. C., CERVANTES, B. M. e GONZALEZ F. A. (2006), “Segmenting wine consumers according to their involvement with appellations of origin”, *Brand Management*, Vol. 13, Nº 4-5, pp. 300-12;
- ROUSSINOV, D. e ZHAO, J. L. (2004), “Text clustering and summary techniques for CRM message management”, *Journal of Enterprise Information Management*, Vol. 17, Nº 6, pp. 424-429;
- RUD, O. P. (2001), “Data Mining Cookbook”, New York, John Wiley & Sons;
- RUDDICK, M. E. (1986), “With life style segmentation banks see customers clearly”, *International Journal of Bank Marketing*, Vol. 18, Nº 12, pp. 26-33;
- RUIZ, M. C. e LASSALA, N. C. (2006), “Segmenting consumer by e-shopping behaviour and online purchase intention”, *Journal of Internet Business*, Nº 3, pp. 1-24;
- RYGIELSKI, C., WANG, J. C. e YEN, D. C. (2002), “Data mining techniques for customer relationship management”, *Technology in Society*, Vol. 24, pp. 483-502;

SARAWAGI, S. e NAGARALU, S. H. (2000), “Data Mining Models as Services on the Internet”, *SIGKDD Explorations*, Vol. 2, N° 1, pp. 24-28;

SEGGERN V. K. e HADAWAY, S. (2004), “Dossier: la segmentation de la clientèle: Bien cerner les profils de clientèle”, *Banque Magazine*, N° 659, pp. 24-25;

SHARMA, A. (1988), “Organizational Decision Making as a Segmentation Base for Industrial Markets”, University of Illinois, Urbana Champaign, IL;

SHARMA, A. e LAMBERT, D. (1994), “Segmentation of Markets Based on Customer Service”, *International Journal of Physical Distribution & Logistics Management*, Vol. 24, N° 4, pp. 50-58;

SHAW M.J., SUBRAMANIAM C., TAN G.W. e WELGE, M. E. (2001), “Knowledge management and data mining for marketing”, *Decision Support Systems*, Vol. 31, pp. 127-137;

SHEKHAR, C., MOIAN, S. e THONNAT, M. (1994), “Towards an intelligent problem solving environment for signal processing”, *Mathematics and computers in Simulation*, Vol. 36, pp.347-359;

SIRGY, M. (1982), “Self-concept in consumer behaviour: a critical review”, *Journal of Consumer Research*, Vol. 9, pp. 287-300;

SMITH, W. (1956), “Product differentiation and market segmentation as alternative marketing strategies”, *Journal of Marketing*, Vol. 21, pp. 3-8;

SOKAL, R. e SNEATH, P. (1963), “Principles of numerical taxonomy”, Freeman, San Francisco;

SOUTAR, G. N, e MCNEIL, M. M. (1991), “A benefit segmentation of financial planning market”, *International Journal of Bank Marketing*, Vol. 9, N° 2, pp. 25-29;

STANLEY, T. O., FORD, J. K. e RICHARDS, S. K. (1985), “Segmentation of bank customers by age”, *International Journal of Bank Marketing*, Vol. 3, N° 3, pp. 56-63;

STEENKAMP, J-B. E. M., e HOFSTEDE, F. (2002), “International market segmentation: Issues and perspectives”, *International Journal of Research in Marketing*, Vol. 19, N° 3, pp. 185-213;

TAM, N. P., CLIQUET G., BORGES, A. e LERAY, F. (2003), “L'opposition entre taille du marché et degré d'homogénéité des segments: Une approche par la logique floue”, *Décisions marketing*, N° 23, pp. 55-69;

TAN, P-N., STEINBACH, M. e KUMAR, V. (2006), “Introduction to Data Mining”, Boston: Addison-Wesley;

TASSEL, D. (2004), “Dossier: la segmentation de la clientèle - Multicanal: comment faire de la segmentation un langage commun”, *Banque Magazine*, N° 659, pp. 30-33;

THELEN S., MOTTNER, S. e BERMAN, B. (2004), “Data mining: On the trail to marketing gold”, *Business Horizons*, Vol. 47, N° 6, pp. 25-32;

TRYON, R. C. (1939), “Cluster analysis”, Edwards Brothers;

TWEDT, D. (1964), “How important to marketing strategy is the heavy user”, *Journal of Marketing*, Vol. 28, pp. 71-72;

VAN RAAIJ, E. M., VERNOOIJ, M. J. A. e VAN TRIES, S. (2003), “The implementation of customer profitability analysis: a case study”, *Industrial Marketing Management*, Vol. 32, N° 7, pp. 573-583;

VAZ, G. N. (2003), “Marketing institucional: o mercado de ideias e imagens”, 2ed São Paulo: Pioneira Thomson Learning;

VERBOVEN, F. (1996), “International Price Discrimination in the European Car Market”, *RAND Journal of Economics, The RAND Corporation*, Vol. 27, N° 2, pp. 240-268;

VERBOVEN, F. (2002), “Quantitative Study to Define the Relevant Market in the Passenger Car Sector”, *Directorate General of Competition of the European Commission*, Catholic University of Leuven;

VIVEROS, M., NEARHOS, J. P. e ROTHMAN, M. J. (1996), “Applying Data Mining Techniques to a Health Insurance Information System”, *Proceedings of the 22th International Conference on Very Large Data Bases*, pp. 286-294;

VOELPEL, S., LEIBOLD, M., TEKIE, E., e VON KROGH, G. (2005), “Escaping the red queen effect in competitive strategy: Sense-testing business models”, *European Management Journal*, Vol. 23, N° 1, pp. 37–49;

WEDEL, M. e KAMAKURA, W. A. (1999), “Market Segmentation – Conceptual and Methodological Foundations”, 2nd edition, Kluwer Academic Publishers;

WEINSTEIN, A. (2004), “Market segmentation”, Homewood, IL: Irwin;

WELLS, W. (1975), “Psychographics: a critical review”, *Journal of Marketing Research*, Vol. 12, pp. 196-213;

WIND, Y. e CARDOZA, R. (1974), “Industrial market segmentation,” *Industrial Marketing Management*, N° 3, pp. 153-66;

WIND, Y. (1978), “Issues and advances in segmentation research”, *Journal of Marketing Research*, Vol. 15, pp. 317-37;

WIND, Y., THOMAS E. e ROBERT, J. (1994), “Segmenting Industrial Markets”, *Advance in Business Marketing and Purchasing*, 6, pp. 59–82;

- XU, R. e WUNSCH II, D. (2005), "Survey of Clustering Algorithms", *IEEE Transactions on Neural Networks*, Vol. 16, N° 3, pp. 645-678;
- YANKELOVICH, D. (1964), "New criteria for market segmentation", *Harvard Business Review*, Vol. 42, March-April, pp. 83-90;
- YAVAS, U. (1988), "Banking behaviour in an Arabian Gulf country: a consumer survey", *International Journal of Bank Marketing*, Vol. 6, N° 5, pp. 40-48;
- YE, N. (2003), "The Handbook of Data mining", Lawrence Erlbaum Associate, New Jersey;
- YELKUR, R. e Da COSTA, M. (2001), "Differential pricing and segmentation on the Internet: the case of hotels", *Management Decision*, Vol. 39, N° 4, pp. 252-262;
- YOUNG, S., OTT, L. e FEIGIN, B. (1978), "Some practical considerations in market segmentation", *Journal of Marketing Research*, Vol. 15, N° 3, pp. 405-12;
- ZAKRZEWSKA, D. e MURLEWSKI, J. (2005), "Clustering Algorithms for Bank Customer Segmentation", *Proceedings of the 5th International Conference on Intelligent Systems Design and Applications*, pp. 197-202;
- ZIKMUND, W. e M. d'AMICO (1993), "Marketing", 4th edition, West Publishing Company, St. Paul, MN;
- ZOTTI, E. (1985), "Thinking psychographically", *Public Relations Journal*, Vol. 41, pp. 26-30;

SÍTIOS ELECTRÓNICOS: (consultados entre Julho de 2008 e Julho de 2009):

Banco de Portugal, www.bportugal.pt, Junho de 2009

Caixa Geral de Depósitos, www.cgd.pt, Julho de 2009

CRISP-DM, www.crisp-dm.org

Santander Totta, www.santandertotta.pt, Julho de 2009

SAS, www.sas.com, Junho de 2009

ANEXOS

ANEXO A

Neste anexo é apresentada de forma resumida a metodologia CRISP-DM.

Metodologia **CRISP-DM**

Na apresentação das fases da metodologia CRISP-DM destacam-se as características mais proeminentes e a documentação produzida. A exposição que se segue tem como base o documento da CRISP-DM (Chapman *et al.*, 2000):

a) Estudo do Negócio

A abordagem ao *negócio* da organização, centra-se na análise dos objectivos do projecto e nos requisitos (funcionais, técnicos, temporais) segundo a perspectiva organizacional.

O conhecimento adquirido neste estudo é posteriormente utilizado para a definição do problema de DM e na concepção do plano preliminar.

A primeira abordagem do processo de DM consiste em estudar a necessidade da realização do próprio projecto de DM, compreender e enquadrar a perspectiva do problema, os objectivos a atingir e descobrir quais os factores mais preponderantes que influenciam os resultados, ou seja, perceber a envolvente do problema a resolver. O estudo do negócio realiza-se pelas seguintes tarefas:

1. *Determinação dos objectivos do negócio* – na fase inicial do projecto é fundamental compreender todos os aspectos que condicionam o negócio como seja, conhecer segundo a perspectiva da organização, os objectivos primários do cliente (*e.g.*, fidelização dos clientes actuais prevendo quando estes estão susceptíveis de abandonar);
2. *Avaliação da situação actual* – determinar com exactidão todos os recursos disponíveis para o projecto (recursos humanos, materiais e financeiros). Realizar um levantamento de todos os requisitos, pressupostos e restrições do projecto, o que inclui um programa de realização, compreensibilidade, qualidade dos

resultados, segurança, aspectos legais e restrições na disponibilidade dos recursos e tecnológicos. deverão igualmente ser identificados todos os riscos, ameaças ou eventos que possam comprometer o projecto e respectivos planos de contingência (acções que previnem o risco). Importa referir ainda a importância da elaboração de uma análise de custos e benefícios para o projecto, onde se compare os custos deste com o potencial benefício para o negócio;

3. *Definição dos objectivos de DM* – descrição dos objectivos de DM e os critérios de sucesso do DM (*e.g.*, classificação, previsão, segmentação). Como exemplo, tendo por base o histórico das compras efectuadas nos últimos anos, o preço dos produtos e a informação demográfica (*e.g.*, idade, rendimentos, cidade, sexo), prever a quantidade que um cliente irá comprar;
4. *Definição do plano para o projecto* – esta tarefa consiste na elaboração de um plano para o projecto que inclua a duração, os recursos, as fases, as sub-fases, as interacções entre os processos, entradas, saídas e dependências. Inclui ainda a elaboração do pressuposto inicial para as ferramentas e técnicas (*e.g.*, requisitos ao nível das ferramentas, dos Sistemas Operativos).

O resultado final do *estudo do negócio* consiste num plano do projecto que inclui a informação acerca do negócio, os seus objectivos e critérios de sucesso, os vários recursos, os requisitos e restrições, os custos e benefícios, os objectivos de DM e os pressupostos das ferramentas e técnicas a utilizar.

b) Estudo dos Dados

A fase de *estudo dos dados*, inicia-se com recolha inicial dos dados e prossegue com a sua análise de forma a identificar problemas de qualidade. Para que se possam aplicar as técnicas de DM aos dados, é necessário ter em conta algumas tarefas, como sejam:

1. *Recolha inicial dos dados* – consiste na aquisição dos dados e da sua compreensão. Desta tarefa resulta uma lista dos dados adquiridos, a sua localização, os métodos de aquisição, problemas e soluções encontradas.

2. *Descrição dos dados* – uma vez recolhidos os dados é necessário descrevê-los, reconhecer o seu formato, o número de registos nas tabelas, identificar os registos e outras características entretanto descobertas.
3. *Exploração os dados* – o resultado desta tarefa consiste numa listagem inicial de hipóteses e o seu impacto no restante projecto. Para uma melhor exploração utilizam-se, por exemplo, gráficos e histogramas, que indicam características dos dados.
4. *Verificação da qualidade dos dados* – realizar um relatório que inclui problemas de qualidade nos dados e possíveis soluções (normalmente dependem directamente dos dados e do conhecimento do negócio).

c) Preparação dos Dados

A fase de *preparação dos dados* envolve todas as actividades associadas à construção do conjunto final de dados, aquele que será usado na ferramenta de modelação, sofrendo inevitavelmente várias optimizações. Esta fase inclui a selecção de tabelas, registos e atributos, bem como a transformação e limpeza dos dados a usar na ferramenta de modelação, as sub-fases são as seguintes:

1. *Seleção de dados* – consiste na escolha dos dados a utilizar na análise. Os critérios para a selecção incluem a relevância dos objectivos de DM e restrições técnicas e de qualidade, como os limites no volume de dados e tipo de dados. No final desta tarefa é ainda realizada uma listagem dos dados incluídos e excluídos e as razões da decisão;
2. *Limpeza de dados* – Complementa a tarefa anterior, existindo várias técnicas que se podem aplicar de forma a optimizar a qualidade dos dados (e.g., a normalização dos dados e tratamento dos dados omissos);
3. *Derivação de dados* – realiza-se pela derivação de novos atributos (e.g. determinar o novo atributo *idade*, a partir da *data de nascimento*), criação de novos registos e transformação dos dados (normalização);

4. *Integração de dados* – obtém-se recorrendo a métodos para a criação de novos registos ou valores, cuja informação é uma combinação de múltiplas tabelas ou registos (*e.g.*, junção e agregação de tabelas ou registos);
5. *Formatação de dados* – a ultima tarefa da preparação dos dados consiste em modificações sintácticas nos dados de modo a que não alterem o seu significado, mas que os tornem utilizáveis pela ferramenta de modelação.

d) Modelação

Esta fase consiste na selecção de várias técnicas de modelação (*e.g.*, árvores de decisão ou redes neuronais artificiais) e os seus parâmetros são ajustados de forma a otimizar os resultados. Normalmente, para o mesmo problema de DM existem várias técnicas disponíveis (*e.g.*, as técnicas árvores de decisão ou redes neuronais artificiais aplicam-se a problemas de classificação), sendo que algumas têm requisitos específicos para a forma como os dados são apresentados, pelo que pode ser necessário voltar à fase da *preparação dos dados*.

Como referido anteriormente, no início do processo, são especificados os problemas e os objectivos do DM, no entanto, apenas nesta fase é que os dados, previamente preparados para a modelação, são utilizados. A escolha das técnicas deve ser cuidadosa de modo a que satisfazer os objectivos de DM. Nesta fase são contempladas as seguintes tarefas:

1. *Seleccção de técnicas de modelação* – a selecção da técnica mais apropriada deve ser realizada tendo em atenção o tipo de problema, as ferramentas e os objectivos do DM;
2. *Definição de uma concepção de teste* – importa antes de construir o modelo, definir um procedimento ou um mecanismo para testar o desempenho do próprio modelo;
3. *Construção do modelo* – uma vez seleccionada a ferramenta de modelação, esta é aplicada ao conjunto de dados preparados anteriormente, permitindo a criação de um ou mais modelos. Os vários parâmetros das ferramentas de modelação

devem ser ajustados e os modelos resultantes devem ser convenientemente interpretados e o seu desempenho explicado;

4. *Revisão do modelo* – a interpretação dos modelos deve ser realizada de acordo com o domínio do conhecimento, critérios de sucesso do projecto de DM e com o mecanismo de teste definido. Na avaliação do sucesso de aplicação do modelo deve ser levada em consideração o impacto dos resultados deste no contexto do negócio.

e) Avaliação

A fase de *avaliação* consiste na validação da utilidade do modelo (ou modelos), na revisão dos passos executados na sua construção e verificação se for atingidos os objectivos do negócio. Esta fase compreende as seguintes tarefas:

1. *Avaliação dos resultados* – determinação se o modelo atingiu os objectivos do negócio (e de DM) e avaliação do modelo quanto a possíveis lacunas;
2. *Revisão do processo* – análise de todas as fases do processo de modo a realçar eventuais actividades esquecidas e/ou que necessitem de ser repetidas;
3. *Determinação dos próximos passos* – o projecto apenas se considera concluído se todos os passos anteriores foram satisfatórios e os resultados cumpriram os objectivos, devendo então passar para a sua fase de implementação. Caso suceda o inverso, é necessário então proceder a uma nova iteração das fases anteriores, utilizando novos parâmetros.

f) Implementação

Na *Implementação*, são empreendidas actividades que conduzem à organização do conhecimento e à sua disponibilização, para que possa ser utilizado no negócio.

A fase de implementação pode ser tão simples como, por exemplo, gerar um relatório ou pode ser tão complexa como integrar os resultados nos sistemas da organização, dependendo dos requisitos. Em muitos casos é o utilizador e não o analista, que executa os passos de implementação, sendo no entanto importante que este

entenda as acções que precisa de executar de forma a fazer uso dos modelos criados. As tarefas envolvidas nesta fase são:

1. *Planeamento da avaliação dos resultados* – define a estratégia para a implementação dos resultados de DM, incluindo os passos e a forma como executar.
2. *Planeamento da monitorização e manutenção* – consiste na definição de estratégia de monitorização e manutenção e é aconselhável sempre que os resultados do DM (modelos), sejam implementados no domínio do problema como parte da rotina do quotidiano. Como retorno da monitorização e manutenção é possível verificar se os modelos são usados correctamente.
3. *Produção de um relatório final* – é a fase de conclusão do projecto de DM, e consiste na elaboração de um relatório final onde devem ser apresentados os pontos mais importantes, a experiência adquirida, e a explicação dos resultados produzidos sobretudo daqueles considerados de maior importância.
4. *Revisão do Projecto* – avaliação dos pontos que foram efectuados de forma correcta, do que correu bem ou que necessita de ser melhorado. Resumo das experiências mais importantes do projecto, torna-se benéfico para projectos futuros e em situações similares referir as armadilhas, aproximações erradas ou como foram seleccionadas as técnicas de DM.

Por concluir, podemos dizer que a metodologia CRISP-DM é muito completa e documentada, uma vez que as suas fases estão devidamente organizadas, estruturadas e definidas, permitindo que o projecto possa ser facilmente compreendido ou revisto.

ANEXO B

Neste anexo é apresentada de forma resumida a metodologia SEMMA.

Metodologia SEMMA

A metodologia SEMMA consiste em cinco fases as quais são seguidamente caracterizadas de acordo com a documentação fornecida pelo instituto *SAS*⁵⁴.

a) Sample – Amostragem

A primeira fase da metodologia SEMMA consiste na realização de uma amostragem, significativa, com a extracção de uma quantidade de dados do universo existente – a amostra deve corresponder a um subconjunto de dados que pertencem ao universo onde cada elemento tem as mesmas hipóteses de ser incluído, mas também deve ser pequena de modo a tornar-se rápida e de fácil manipulação.

A realização do processo de amostragem traduz-se numa optimização dos custos, da rentabilidade e do desempenho das etapas seguintes, dado o facto de a manipulação de uma amostra ser mais rápida e fácil do que manipular todo o universo de dados disponíveis.

O desenvolvimento de todo o processo de DM a partir de uma amostra representativa reduz drasticamente o volume e o tempo de processamento necessário para tirar informação crucial para o negócio. Neste contexto, se o universo de dados tiver um determinado padrão ou tendência bastante determinado, estes estão patentes na amostra, caso contrário, o padrão ou tendência for irrelevante, ao ponto de não ser detectado na amostra, também não será importante para o universo de dados (SAS, 2009).

b) Explore – Exploração

⁵⁴ www.sas.com.

Uma vez realizado o processo de amostragem, a primeira abordagem realizada sobre os dados consiste em explorá-los visualmente ou numericamente (*e.g.*, gráficos de distribuição e dispersão, histogramas) permitindo em alguns caso detectar as tendências ou agrupamentos inerentes nos dados. A *exploração* ajuda a refinar o processo de descoberta.

c) Modify - Modificação

A fase da *modificação* concentra todas as transformações necessárias com base nos resultados da etapa de exploração. As transformações realizadas podem ser de inclusão de informação (*e.g.*, agrupamento de subgrupos significativos de dados), selecção ou introdução de novas variáveis, de forma a obter-se as variáveis mais significativas. O objectivo desta fase consiste em criar, seleccionar e transformar as variáveis para o processo de construção do modelo, preparando os dados para a próxima etapa.

d) Model - Modelação

Uma vez preparados os dados, é possível então prosseguir para a fase de aplicação de algoritmos - *modelação*. É nesta fase que se definem as técnicas de construção de modelos de DM, onde se incluem as técnicas de aprendizagem automática (*e.g.*, árvores de decisão ou redes neuronais artificiais) e modelos estatísticos (*e.g.*, regressão linear). Para a selecção da técnica é necessário levar em consideração que cada modelo tem propriedades e características singulares dependentes dos dados e adequados a situações específicas de DM (*e.g.*, as redes neuronais artificiais alcançam melhores resultados com dados com relacionamentos complexos e não lineares). A etapa de modelação tem como objectivo seleccionar as técnicas de construção de modelos de forma a prever com confiança os resultados desejados (SAS, 2009).

e) Assessment – Avaliação

A etapa final da metodologia SEMMA consiste na *avaliação* do modelo de forma a aferir o seu desempenho. Geralmente, a fase de avaliação de um modelo corresponde à aplicação deste a uma amostra de dados seleccionada para este fim (conjunto de teste).

Se o modelo for válido este deve funcionar tão bem como na amostra que serviu de base à sua construção. A etapa de avaliação tem como objectivo aplicar o modelo à amostra de dados e verificar a seu desempenho, de forma a proceder a ajustes se necessário (SAS, 2009).

ANEXO C

Neste anexo é apresentada de forma resumida os índices de similaridade.

Tabela C.1 – Índices de similaridades (Referências e Símbolos)

Nº	Símbolo	Formula	Intervalo	Introduzido por
1	R	$\frac{a+d}{a+b+c+d}$	[0,1]	Sokal e Michener (1958), Rand (1971)
2	H	$\frac{(a+d)-(b+c)}{a+b+c+d}$	[-1,1]	Hamann (1961), Hubert (1977)
3	CZ	$\frac{2a}{2a+b+c}$	[0,1]	Czekanowski (1932), Dice (1945), Gower e Legendre (1986)
4	K	$\frac{1}{2} \left(\frac{a}{a+b} + \frac{a}{a+c} \right)$	[0,1]	Kulczynski (1972)
5	MC	$\frac{a^2 - bc}{(a+b)(a+c)}$	[-1,1]	McConnaughey (1964)
6	PE	$\frac{ad - bc}{(a+c)(b+d)}$	[-1,1]	Peirce (1884)
7	FM	$\frac{a}{\sqrt{(a+b)(a+c)}}$	[0,1]	Fowlkes e Mallows (1983), Ochiai (1957)
8	W1	$\frac{a}{a+b}$	[0,1]	Wallace (1) (1983)
9	W2	$\frac{a}{a+c}$	[0,1]	Wallace (2) (1983)
10	Γ	$\frac{ad - bc}{\sqrt{(a+b)(a+c)(c+d)(b+d)}}$	[-1,1]	Gamma
11	SS1	$\frac{1}{4} \left(\frac{a}{a+b} + \frac{a}{a+c} + \frac{d}{d+b} + \frac{d}{d+c} \right)$	[0,1]	Sokal e Sneath (1963)
12	B1	$\frac{\binom{m}{2} - \binom{m}{2} (b+c) + (b-c)^2}{\binom{m}{2}}$	[0,1]	Baulieu (1989)
13	RR	$\frac{a}{a+b+c+d}$	[0,1]	Russel e Rao (1940)
14	FMG	$\frac{\sqrt{(a+b)(a+c)} - \frac{1}{2\sqrt{a+b}}}{ad - bc}$	[-½,1]	Fager e McGowan (1963)
15	P	$\frac{ad - bc}{(a+b)(a+c)(c+d)(b+d)}$	[-1,1]	Pearson
16	B2	$\frac{ad - bc}{\binom{m}{2}}$	[-¼, ¼]	Baulieu (1989)
17	J	$\frac{a}{a+b+c}$	[0,1]	Jaccard (1908)
18	SS2	$\frac{a}{a+2(b+c)}$	[0,1]	Sokal e Sneath (1963)
19	SS3	$\frac{ad}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$	[0,1]	Sokal e Sneath (1963), Ochiai (1957)
20	GL	$\frac{a+d}{a+\frac{1}{2}(b+c)+d}$	[0,1]	Gower e Legendre (1986), Sokal e Sneath (1963)
21	RT	$\frac{a+d}{a+2(b+c)+d}$	[0,1]	Rogers e Tanimoto (1960)
22	GK	$\frac{ad - bc}{ad + bc}$	[-1,1]	Goodman e Kruskal (1954), Yule (1927)

Adaptado de Albatineh *et al.* (2006)

ANEXO D

Neste anexo é apresentado o resultado da selecção das segmentações das abordagens realizadas neste trabalho.

A seguir vamos apresentar, em primeiro lugar, os resultados obtidos com a primeira abordagem (selecção “simples”). Em segundo lugar, apresentamos os resultados obtidos com a segunda abordagem (selecção de *clusterings* usando índices de similaridade).

Na primeira abordagem, efectuou-se, em primeiro lugar, uma comparação dos conjuntos (de pares) com o mesmo número de clusters (ver tabela apresentada abaixo).

Tabela D.1. – Comparação de *clusterings* com o mesmo número de clusters

3x3	5x5	10x10
H3F - H3P	H5F - H5P	H10F - H10P
C3F - C3P	C5F - C5P	C10F - C10P
K3F - K3P	K5F - K5P	K10F - K10P
W3F - W3P	W5F - W5P	W10F - W10P
H3F - C3F	H5F - C5F	H10F - C10F
H3F - C3P	H5F - C5P	H10F - C10P
H3P - C3F	H5P - C5F	H10P - C10F
H3P - C3P	H5P - C5P	H10P - C10P
C3P - K3P	C5P - K5P	C10P - K10P
C3F - K3F	C5F - K5F	C10F - K10F
C3P - K3F	C5P - K5F	C10P - K10F
C3F - K3P	C5F - K5P	C10F - K10P
K3P - H3P	K5P - H5P	K10P - H10P
K3F - H3F	K5F - H5F	K10F - H10F
K3P - H3F	K5P - H5F	K10P - H10F
K3F - H3P	K5F - H5P	K10F - H10P
W3P - H3P	W5P - H5P	W10P - H10P
W3F - H3F	W5F - H5F	W10F - H10F
W3P - H3F	W5P - H5F	W10P - H10F
W3F - H3P	W5F - H5P	W10F - H10P
W3P - K3P	W5P - K5P	W10P - K10P
W3F - K3F	W5F - K5F	W10F - K10F
W3P - K3F	W5P - K5F	W10P - K10F
W3F - K3P	W5F - K5P	W10F - K10P
C3P - W3P	C5P - W5P	C10P - W10P
C3F - W3F	C5F - W5F	C10F - W10F
C3P - W3F	C5P - W5F	C10P - W10F
C3F - W3P	C5F - W5P	C10F - W10P

Em segundo lugar, usámos dois critérios para a (pré-) selecção dos *clusterings* que são: o maior número de células preenchidas (NCP) e o valor mais elevado na matriz de confusão (MaxM), descritos no capítulo V. De acordo com os resultados obtidos, seleccionámos os seguintes *clusterings* (ver tabela D.2.):

Tabela D.2. – (Pré-)Seleccção dos *clusterings* na primeira etapa da metodologia de selecção simples

Cluster de 3	Cluster de 5	Cluster de 10
K3P	K5P	K10P
K3F	H5P	W10F
W3F	W5F	W10P
W3P	W5P	

Após a análise de todos os conjuntos de pares possíveis (ver tabela D.3.) foram seleccionadas as segmentações seguintes: K3P, K5P, H5P, K10P e W10P usando os critérios o NCP e o MaxM.

Tabela D.3. – Comparação dos *clusterings* com diferente número de clusters (NC)

3x5	3x10	5x10
K3F - K5P	K3P - W10F	K5P - W10F
W3F - K5P	K3F - W10F	W5P - W10F
K3P - K5P	W3F - W10F	H5P - W10F
W3P - K5P	W3P - W10F	W5F - W10F
K3F - W5P	K3P - K10P	K5P - K10P
W3F - W5P	K3F - K10P	W5P - K10P
K3P - W5P	W3F - K10P	H5P - K10P
W3P - W5P	W5P - K10P	W5F - K10P
K3F - H5P	K3P - W10P	K5P - W10P
W3F - H5P	K3F - W10P	W5P - W10P
K3P - H5P	W3F - W10P	H5P - W10P
W3P - H5P	W3P - W10P	W5F - W10P
K3F - W5F		
W3F - W5F		
K3P - W5F		
W3P - W5F		

Para a primeira abordagem, seleccionámos as segmentações seguintes (ver tabela D.4.):

Tabela D.4. – Seleccção das segmentações para a primeira abordagem

3x5	H5P, K3P, W3F
3x10	K10P, K3P
5x10	K10P, K5P, W10P

Convém referir que, nesta abordagem, optámos por escolher seis segmentações (*clustering*) a fim de simplificar a análise: duas segmentações com 3 clusters; duas com 5 clusters; e duas segmentações com 10 clusters.

Em relação à segunda abordagem, comparámos todos os pares com o mesmo número de clusters usando o índice R (Rand, 1971).

Os resultados obtidos foram apresentados em tabelas com o respectivo número de clusters (NC):

Tabela D.5. – Resultados com NC=3

R	H3F	C3F	K3F	W3F	H3P	C3P	K3P	W3P
H3F		1,00	1,00	0,55	0,99	0,96	0,45	0,65
C3F	1,00		1,00	0,55	0,99	0,96	0,45	0,65
K3F	1,00	1,00		0,55	0,99	0,96	0,45	0,65
W3F	0,55	0,55	0,55		0,55	0,55	0,49	0,50
H3P	0,99	0,99	0,99	0,55		0,97	0,45	0,64
C3P	0,96	0,96	0,96	0,55	0,97		0,47	0,63
K3P	0,45	0,45	0,45	0,49	0,45	0,47		0,47
W3P	0,65	0,65	0,65	0,50	0,64	0,63	0,47	

Podemos constatar por exemplo que os *clusterings* H3F e C3F são similares, enquanto que o W3P e W3F são diferentes (ver tabela D.5.). Após a análise dos resultados obtidos no NC3, seleccionámos os *clusterings* K3P, H3F, W3F e W3P.

Tabela D.6. – Resultados com NC=5

R	H5F	C5F	K5F	W5F	H5P	C5P	K5P	W5P
H5F		1,00	0,99	0,55	0,88	0,87	0,36	0,38
C5F	1,00		0,99	0,55	0,88	0,87	0,36	0,38
K5F	0,99	0,99		0,56	0,87	0,87	0,37	0,39
W5F	0,55	0,55	0,56		0,52	0,53	0,48	0,48
H5P	0,88	0,88	0,87	0,52		0,93	0,42	0,45
C5P	0,87	0,87	0,87	0,53	0,93		0,45	0,46
K5P	0,36	0,36	0,37	0,48	0,42	0,45		0,64
W5P	0,38	0,38	0,39	0,48	0,45	0,46	0,64	

Relativamente aos resultados obtidos com NC=5 (ver tabela D.6.), verificámos neste caso que por exemplo que C5F e H5F são similares. Os *clusterings* seleccionados no NC5 foram C5F, W5F, K5P e W5P.

Tabela D.7. – Resultados com NC=10

R	H10F	C10F	K10F	W10F	H10P	C10P	K10P	W10P
H10F		1,00	0,52	0,51	0,53	0,66	0,24	0,19
C10F	1,00		0,52	0,51	0,53	0,66	0,24	0,19
K10F	0,52	0,52		0,82	0,47	0,48	0,48	0,49
W10F	0,51	0,51	0,82		0,48	0,47	0,49	0,49
H10P	0,53	0,53	0,47	0,48		0,82	0,66	0,57
C10P	0,66	0,66	0,48	0,47	0,82		0,54	0,47
K10P	0,24	0,24	0,48	0,49	0,66	0,54		0,78
W10P	0,19	0,19	0,49	0,49	0,57	0,47	0,78	

Finalmente no NC=10 (ver tabela D.7.), verificámos que o C10F e o H10F são similares, e o C10P e o H10P são muito similares por exemplo. Os *clusterings* seleccionados no NC10 foram o C10P, H10F, K10F, K10P e o W10P.

A partir dos resultados obtidos com a comparação dos conjuntos de pares com o mesmo número de clusters. Comparámos 5x3, 10x3 e 10x5.

Seleccionámos as segmentações H3F, K3P, K5P, W5F, K10P e W10P, como se pode ver nas tabelas seguintes:

Tabela D.8. – Resultados da metodologia usando índices de similaridade

5x3				
R	H3F	K3P	W3F	W3P
C5F	1,00	0,45	0,55	0,65
W5F	0,55	0,50	1,00	0,73
K5P	0,36	0,36	0,48	0,43
W5P	0,38	0,60	0,48	0,53

10x3				
R	H3F	K3P	W3F	W3P
C10P	0,66	0,60	0,50	0,54
H10F	1,00	0,45	0,55	0,65
K10F	0,52	0,49	0,81	0,49
K10P	0,24	0,74	0,47	0,45
W10P	0,18	0,61	0,47	0,53

10x5				
R	C5F	W5F	K5P	W5P
C10P	0,66	0,50	0,58	0,54
H10F	1,00	0,55	0,36	0,38
K10F	0,52	0,81	0,49	0,48
K10P	0,24	0,47	0,83	0,69
W10P	0,18	0,47	0,68	0,80

Por exemplo, no caso 5x3, escolhemos as segmentações K5P, H3F e K3P por terem os valores mais baixos, depois não escolhemos C5F nem W5P por terem um valor muito elevado.

Em relação ao caso 10x5, escolhemos a segmentação W5F, porque tinha dois valores razoáveis. Depois escolhemos a W10P, porque tinha os valores baixos relativamente às quatro segmentações escolhidas nos passos anteriores.

Convém referir que a título de simplificação decidimos escolher seis segmentações: duas segmentações com 3 clusters; duas com 5; e duas segmentações com 10 clusters.

ANEXO E

Neste anexo é apresentada as tabelas das segmentações seleccionadas com o alinhamento dos cartões bancários do Santander Totta e da Caixa Geral de Depósitos

Tabela E.1. – Alinhamento dos cartões do S.T. e da C.G.D. com as Segmentações W3F, K3P

Bancos	Cartões	W3F			K3P		
		S1	S2	S3	S1	S2	S3
S.T.	Platinum						
	GOLD						
	Premium Travel						
	Light						
	5%						
	10.10 Tsi						
	Electron						
	Premier						
C.G.D.	Recarregáveis						
	Caixa Fã						
	Caixautomática Electron / Maestro						
	Farmácias Portuguesas						
	MTV						
	Megacartão Jovem						
	Caixa Woman						
	Caixa Carbono Zero						
	Caixa Gold						
	Caixa Classic						
	Impar						
	LEVE						
	Miles & More						
	Soma						
	Lol (Júnior / Lol / Pro)						

Legendas: **amarelo** – possível **verde** – segmento coberto pelo cartão

Tabela E.2. – Alinhamento dos cartões do S.T. e da C.G.D. com as Segmentações W5F, K5P e H5P

Bancos	Cartões	W5F					K5P					H5P				
		S1	S2	S3	S4	S5	S1	S2	S3	S4	S5	S1	S2	S3	S4	S5
S.T.	Platinum															
	GOLD															
	Premium Travel															
	Light															
	5%															
	10.10 Tsi															
	Electron															
	Premier															
C.G.D.	Recarregáveis															
	Caixa Fã															
	Caixautomática Electron / Maestro															
	Farmácias Portuguesas															
	MTV															
	Megacartão Jovem															
	Caixa Woman															
	Caixa Carbono Zero															
	Caixa Gold															
	Caixa Classic															
	Impar															
	LEVE															
	Miles & More															
	Soma															
	Lol (Júnior / Lol / Pro)															

Legendas: **amarelo** – possível **verde** – segmento coberto pelo cartão

Tabela E.3. – Alinhamento dos cartões do S.T. e da C.G.D. com as Segmentações K10P e W10P

Bancos	Cartões	K10P										W10P									
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S.T.	Platinum				Amarelo											Amarelo					
	GOLD		Amarelo																Verde	Verde	Verde
	Premium Travel				Verde											Verde					
	Light	Amarelo		Verde			Amarelo		Amarelo	Verde					Amarelo						
	5%	Verde	Verde				Verde	Verde	Verde	Amarelo	Verde	Verde	Verde	Verde	Verde		Verde	Amarelo			
	10.10 Tsi					Verde										Verde					
	Electron	Verde		Verde		Amarelo	Verde	Amarelo		Verde		Amarelo	Verde	Verde	Verde	Verde	Verde	Verde	Amarelo		
	Premier										Verde		Verde	Amarelo		Amarelo					
	Recarregáveis	Amarelo											Amarelo						Amarelo		
C.G.D.	Caixa Fã								Verde	Verde											
	Caixautomática Electron / Maestro	Amarelo		Verde		Amarelo	Amarelo	Amarelo		Verde		Amarelo	Verde	Verde	Verde	Verde	Verde	Amarelo	Amarelo		
	Farmácias Portuguesas							XXX				XXX									
	MTV								Amarelo												
	Megacartão Jovem								Amarelo	Amarelo					Amarelo						
	Caixa Woman	Verde	Verde				Verde	Verde		Verde	Verde	Verde	Verde	Verde	Verde			Verde	Amarelo	Amarelo	Verde
	Caixa Carbono Zero																				
	Caixa Gold		Amarelo				Verde												Verde	Verde	Verde
	Caixa Classic	Verde		Amarelo							Verde		Verde	Verde	Verde			Amarelo	Amarelo	Amarelo	Amarelo
	Impar			Verde																	
	LEVE																				
	Miles & More				Verde											Verde					
	Soma					Verde											Verde				
	Lol (Júnior / Lol / Pro)								Amarelo				Amarelo								

Legendas: **amarelo** – possível **verde** – segmento coberto pelo cartão

XXX – possível (se pressuposto incluído na campanha do cartão)