



NOVA

IMS

Information
Management
School

MGI

Mestrado em Gestão de Informação

Master Program in Information Management

Retalho Omnicanal

Segmentação de Clientes que compraram em lojas físicas após serem acionados por email marketing

Daniel Bassoli

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Gestão de Informação

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

RETALHO OMNICAL

SEGMENTAÇÃO DE CLIENTES QUE COMPRARAM EM LOJAS FÍSICAS APÓS SEREM ACIONADOS POR EMAIL MARKETING

por

Daniel Bassoli

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Gestão de Informação, com Especialização em Business Intelligence

Orientador: Prof. Dr. Rui Gonçalves

Coorientador: Luís Miguel Santos

Maio, 2017

DEDICATÓRIA

Para o meu primeiro filho, ou filha, que neste momento é apenas uma pequena semente no ventre da minha esposa. Aguardo-te para juntos exploramos o mundo.

AGRADECIMENTOS

Agradeço a todos os portugueses que conviveram comigo durante esta incrível jornada para o Mestrado.

A alguns em especial: à Dona Manuela da Silva Garcia (in memoriam), ao Orientador desta Dissertação, Professor Doutor Rui Gonçalves, ao Coorientador desta Dissertação e Gerente de Marketing Digital da Fnac, Luís Miguel Santos, à Gerente de CRM da Fnac, Inês Pereira, à CMO da Fnac, Inês Condeço, e ao Professor Doutor Fernando Bação.

RESUMO

Há uma tendência no Retalho Europeu e Mundial para a implementação da estratégia Omnicanal, ou seja, vários canais sendo usados de forma convergente e complementar, tanto para as comunicações de marketing como para otimizar atividades operacionais. Tendo em vista que a estratégia Omnicanal é decorrência direta do avanço técnico e da disseminação das Tecnologias da Informação, principalmente no contexto do Cloud Computing e do Big Data, torna-se então viável que, a partir da imensa disponibilidade dos dados para a recolha, inerente a este ambiente digital, todos os esforços e atividades que utilizam estas Tecnologias possam ser mensuradas, não apenas com indicadores de performance mas com quaisquer outros indicadores que façam sentido para o negócio. Dentro da estratégia Omnicanal do Retalho esta possibilidade permite analisar por exemplo a influência que um canal possa ter tido na conversão em vendas dentro de uma campanha e mapear a composição dos grupos de clientes que chegaram à conversão. Devido à natureza digital da estratégia, este tipo de análise é comum às campanhas das lojas online. Quando se trata das lojas físicas, no entanto, este tipo de análise ainda não está disseminada. A partir da Análise de Clusters de clientes de uma rede Retalhista multinacional que, dentro de uma campanha foram acionados por um canal digital (Email Marketing), e chegaram à conversão nas lojas físicas, foi proposto um modelo para a segmentação de clientes para a estratégia Omnicanal.

PALAVRAS-CHAVE

Retalho Omnicanal; Clustering; Segmentação; Email Marketing.

ABSTRACT

There is a trend in the European and Global Retailing for the implementation of the Omnichannel strategy, that is, several channels being used in a convergent and complementary way, both for marketing communications and to improve operational activities. Considering that the Omnichannel strategy is a direct result of the technical advancement and dissemination of Information Technologies, especially in the context of Cloud Computing and Big Data, it is then feasible that, given the immense availability of data for the collection, inherent to this digital environment, all the efforts and activities that use these technologies can be measured, not only with performance indicators but with any other indicator that make sense for the business. Within the Omnichannel Retailing strategy, this possibility allows us to analyze, for example, the influence that a channel may have had on the conversion to sales within a campaign, and to map the composition of the client groups that came to conversion. Due to the digital nature of the strategy, this type of analysis is common to online store campaigns. When it comes to physical stores, however, this type of analysis is not yet widespread. Based on the Clustering Analysis of customers from a global Retailer that within a campaign were triggered by a digital channel (email marketing), and came to the conversion at the physical stores, a customer segmentation model was proposed for the Omnichannel strategy.

KEYWORDS

Omnichannel Retailing; Clustering; Segmentation; Email Marketing.

ÍNDICE

1. Introdução	1
1.1. Contexto	1
1.2. Fnac.....	3
2. Relevância do tema	6
3. Objetivos.....	8
3.1. Objetivo Principal	8
3.2. Objetivos Específicos	8
3.2.1. Recolha dos Dados	8
3.2.2. Preparação e Pré-processamento dos Dados	9
3.2.3. Segmentação	9
3.2.4. Avaliação	9
4. Revisão da Literatura	10
5. Metodologia	11
5.1. Recolha dos Dados	11
5.2. Preparação e Pré-processamento dos dados.....	14
5.3. Nova Preparação e Pré-processamento dos dados	17
5.4. Segmentação	21
6. Resultados e Discussão.....	23
6.1. Avaliação.....	23
6.2. Profiling.....	24
7. Conclusões.....	28
8. Limitações e Recomendações para Trabalhos Futuros	30
9. Bibliografia.....	31
10. Anexos	33
10.1. Anexo A: Exemplo de Email Marketing DNA Fnac.....	33
10.2. Anexo B: Diagrama Analítico do projeto em SAS Enterprise Miner	34

ÍNDICE DE FIGURAS

Figura 1 - Mapa das lojas Fnac em Portugal	4
Figura 2 - Peça publicitária dos DNA Fnac de outubro de 2016	5
Figura 3 - Omni-channel: Different Channel, Same Message	6
Figura 4 - Nodes File Import e Save Data	13
Figura 5 - 4 Clusters sem a variável sex_cliente.....	21
Figura 6 - Mapa obtido por Multidimensional Scaling.....	23
Figura 7 - Peça de Email Marketing DNA Fnac de março de 2017.....	31
Figura 8 - Diagrama Analítico do modelo de Segmentação.....	32

ÍNDICE DE GRÁFICOS

Gráfico 1 - Distribuição da variável idade	18
Gráfico 2 - Distribuição da variável Compras_Ano	18
Gráfico 3 - Distribuição da variável CA_DNA	19
Gráfico 4 - Distribuição da variável CA_ANO	19
Gráfico 5 - Distribuição da variável REGIAO_Loja_DNA.....	19
Gráfico 6 - Distribuição da variável REGIAO_Loja_Influ.....	21
Gráfico 7 - Dimensões dos Clusters. SAS Enterprise Miner 9.4	24
Gráfico 8 - Variable Worth do Cluster 1	25
Gráfico 9 - Variable Worth do Cluster 2	25
Gráfico 10 - Histograma da variável data_venda do Cluster 2	26
Gráfico 11 - Variable Worth do Cluster 3	26
Gráfico 12 - Histograma da variável data_venda do Cluster 3	26
Gráfico 13 - Variable Worth do Cluster 3	27
Gráfico 14 - Histograma da variável data_venda do Cluster 4	27

ÍNDICE DE TABELAS

Tabela 1 - Metadata Original.....	13
Tabela 2 - Códigos das Lojas Fnac Portugal.....	14
Tabela 3 - Registos da tabela de interação do Email Marketing da Campanha DNA	14
Tabela 4 - Metadata com os Roles e Levels	16
Tabela 5 - Estatísticas das variáveis de Classe (A).....	16
Tabela 6 - Estatísticas das variáveis de Classe (B).....	16
Tabela 7 - Estatísticas das variáveis Intervalares	17
Tabela 8 - Nova Metadata com os Roles e Levels.....	21
Tabela 9 - Estatísticas das variáveis de Classe	21
Tabela 10 - Matriz de distância entre os Clusters	24

LISTA DE SIGLAS E ABREVIATURAS

APP	aplicativo para plataformas móveis
BI	Business Intelligence, em tradução livre “Inteligência de Negócios”
CRM	Customer Relationship Management, em tradução livre “Gestão do Relacionamento com os Clientes”
DNA	Dias e Noites do Aderente
KPI	Key Performance Indicator, em tradução livre “Indicador-chave de Performance”
NUTS II	Nomenclatura de Unidades Territoriais para Fins Estatísticos de Nível II
ROI	Return on Investment, em tradução livre “Retorno sobre o Investimento”
RFM	Recência, Frequência, e Valor Monetário
SAS	Statistical Analysis System, em tradução livre “Sistema de Análises Estatísticas”
SAAS	Software as a Service, em tradução livre “Software como um Serviço”
SEM	Search Engine Marketing, em tradução livre “Marketing em Mecanismos de Busca”
SEMMA	Sample, Explore, Modify, Model, Assess, em tradução livre “Tirar Amostra, Explorar, Modificar, Modelar, Avaliar”
SMS	Short Message Service, em tradução livre “Serviço de Mensagens Curtas”
SQL	Structured Query Language, em tradução livre “Linguagem de Consulta Estruturada”
TI	Tecnologias da Informação

1. INTRODUÇÃO

1.1. CONTEXTO

O Retalho tradicional Europeu e Global passa por um momento de transição dos modelos “Brick-and-mortar” (com operações exclusivamente em lojas físicas), “Click-and-mortar” (que além das lojas físicas operam também com websites transacionais), ou “Pure-Digital” (também conhecido como E-Commerce) para Omnicanal (“Omni”, do latim “todos”, “universal”), ou seja, atuam em vários canais de contato com os clientes, não apenas para fins de comunicação, mas também como plataformas de venda.

A nível tecnológico, este fenómeno acontece como consequência do impacto qualitativo que as TI – Tecnologias da Informação vêm tendo sobre os vários sectores de negócios há mais de duas décadas (Chan, 2000). A implementação do modelo Omnicanal no Retalho só é possível através de sistemas que permitem o gerenciamento da informação em todos os canais.

Um dos resultados do impacto das TI nos negócios foi o advento de soluções que permitiram a otimização da Gestão do Relacionamento com os Clientes – também chamada CRM (Customer Relationship Management). Os sistemas de CRM, desenvolvidos inicialmente apenas por gigantes como SAP, Oracle, Microsoft, entre outros, eram restritos a empresas de grande porte, principalmente da Banca, das grandes redes do Retalho e do próprio sector de Tecnologia, devido aos custos elevados de implementação, principalmente com a infraestrutura.

Com a consolidação mercadológica do Cloud Computing e os SaaS – Software as a Service, surgiram diversas soluções em CRM de baixo custo, tanto monetário como computacional, o que permitiu que empresas de pequeno e médio porte também pudessem adotar processos do CRM que, vale ressaltar, não se trata apenas de uma solução tecnológica específica, mas de uma filosofia organizacional. Kamakura et al. (2006) define CRM como “o processo de coleta e análise de informações de uma empresa em relação a interações com o cliente a fim de aumentar os valores dos clientes para a empresa” (tradução livre).

O CRM é usualmente classificado em Operacional ou Analítico. Alguns autores como Gneiser (2000) ainda sugerem uma terceira categoria, o CRM Colaborativo – associado ao uso dos processos do CRM entre os stakeholders (em tradução livre, “partes interessadas”) de uma organização. De acordo com Torggler (2008), o CRM Operacional é a dimensão tecnológica, resumida à própria infraestrutura, e tem como ponto-chave a automatização, não apenas das atividades específicas do Marketing, como Gestão de Campanhas e emissão de Relatórios, mas também das outras unidades de negócios interdependentes como Vendas e Serviço Ao Cliente.

Com a estrutura propícia para a recolha dos dados e a possibilidade de gerar informação pelo CRM Operacional, surge então outra frente de atuação, a Analítica. Para Kamakura et al. (2006), o uso de dados dos clientes para a criação dos “modelos de escolha”, e usá-los para o aumento de receita, constitui o núcleo do CRM Analítico. Seus sistemas são focados na recolha sistemática, avaliação e análise de todos os dados referentes aos Clientes (Gneiser, 2010). Para isto faz uso de diversas ferramentas analíticas, como por exemplo o Data Mining que, no contexto do CRM, geram

segmentações refinadas, modelos para geração de Cross-selling (em tradução livre, “vendas cruzadas”) e Up-selling (em tradução livre, “vendas mais rentáveis”), previsão das taxas de Churn (em tradução livre, “agitação”; neste contexto, “evasão”), entre outras análises que possam dar suporte às tomadas de decisão.

O Data Mining ou Mineração de Dados é definido por Berry e Linoff (2004) como “a exploração e análise de grandes quantidades de dados a fim de descobrir padrões e regras significativas”. Há muitas definições para Data Mining devido à universalidade da aplicação de suas técnicas, empregadas em vários campos da ciência, desde as próprias Ciências Computacionais, até às Ciências Naturais, como a Biologia.

Para complementar a definição Data Mining ou Mineração de Berry e Linoff (2004), a definição de Hand et al. (2001) é um pouco mais abrangente, mas apresenta o mesmo viés técnico: “Data Mining é a análise de conjuntos de dados observacionais (muitas vezes grandes) para encontrar relacionamentos insuspeitos e resumir os dados em novas formas que são compreensíveis e úteis para o proprietário dos dados” (em tradução livre).

O Data Mining está associado a modelos matemáticos computacionais que visam a representar a realidade a partir de uma amostra de dados observacionais, ou seja, que já tenham sido recolhidos para alguma finalidade, ao contrário dos dados experimentais ou utilizados pela Estatística para resolver um problema específico. Ele é produto da contribuição de diversas disciplinas como a própria Estatística, Learning Machine (em tradução livre, “Aprendizado de máquina), Tecnologias de Bases de Dados, entre outras.

Dentre as várias aplicações comerciais de suas técnicas, o Data Mining permite que a organização melhore seu desempenho em Marketing através da melhor compreensão de sua base de Clientes. Como ferramenta para Gestão de Campanhas, o Data Mining gera informação que permite, por exemplo determinar quais plataformas de comunicação devem ser acionadas (Web, SMS, Email, Eventos, Social Networks, etc.), quais segmentos de Clientes devem ser contatados, o tipo de linguagem a ser empregado, os preços e produtos a serem destacados, entre outras aplicações.

Dentro da estratégia Omnicanal do Retalho, definida por Rigby (2011) como “uma experiência de vendas integrada que mescla as vantagens das lojas físicas com a experiência rica em informações das compras on-line”, estas aplicações do Data Mining são vitais no contexto atual do *Big Data* (exposição crescente aos dados estruturados e não-estruturados), já que fornece subsídios que permitem manter a mesma informação, preservadas as peculiaridades, através dos canais – um dos fatores que garantem ao Cliente uma experiência de compra satisfatória (SHANKAR ET AL., 2011).

Pela perspectiva Analítica do Data Mining, no âmbito da estratégia e do contexto citados, faz-se necessária a implementação de KPIs – Indicadores-chave de Performance, de acordo com as necessidades de negócio e que façam sentido aos vários canais. As Analíticas devem combinar os dados offline e online.

1.2. FNAC

Uma das grandes redes globais do Retalho que faz uso da estratégia Omnicanal é a Fnac. De acordo com o periódico online francês LSA Commerce & Consommation (2016), “as vendas Omnicanal representaram 46% dos pedidos efetuados pelo E-commerce Fnac.com em 2015”. Estas transações aconteceram na loja online, mas os produtos foram levantados em lojas físicas.

A Fnac, acrônimo do francês *Fédération nationale d'achats des cadres*, em tradução livre "Federação Nacional de Compras para Executivos", foi fundada em França no ano de 1954 por André Essel e Max Théret e, como o nome sugere, nasceu como um clube exclusivo, para a compra de produtos com descontos. Os membros eram informados das ofertas através da revista Contact, também de propriedade de Essel e Théret. Teve sua primeira loja inaugurada em 31 de julho de 1954 em Paris, e desde início já possuía um viés cultural, tendo uma sala reservada para a exposição de fotografias.

Durante as décadas de 60 e 70 apresentou grande expansão, comercializando câmaras, rádios, sistemas de som e outros dispositivos eletrônicos. Em 1966 a Fnac passa a ser aberta ao público geral e três anos mais tarde inaugura sua segunda loja em Paris. Durante a década de 1970 passa a comercializar livros e vendê-los em até 80% do preço praticado pelo Retalho, o que levou à uma forte reação das Editoras, Escritores e outros atores da cadeia produtiva do Livro. No mesmo período foi criado o Forum, espaço dedicado a eventos culturais, presente nas maiores lojas da rede até hoje.

Na década de 1980, já com outras lojas inauguradas e uma rede estabelecida, a Fnac torna-se uma empresa de capital aberto e dá início ao seu processo de internacionalização, e começar por Bélgica, seguida por Espanha e Portugal, já na década de 1990. Hoje a Fnac opera além destes países no Brasil, Suíça, e sob franquia em Marrocos, Qatar, e Costa do Marfim. A rede é composta atualmente por 199 lojas, incluído as lojas online. De acordo com o website institucional do Grupo Fnac (2016), a empresa obteve em 2015 uma receita de € 3,876 bilhões, e lucro operacional de € 85 milhões. Das vendas de 2015, 56,1% foram de Produtos Eletrônicos, 38,8% de Produtos Editoriais (livros, jogos, entre outros), e 5,1% de Serviços.

Em Portugal a Fnac iniciou suas operações no dia 28 de fevereiro de 1998, com a loja do Centro Comercial Colombo. Tem hoje uma rede composta por 22 lojas físicas (Aeroporto de Lisboa; Amoreiras; Alfragide; Algarve Shopping; Almada; Braga; Cascais Shopping; Chiado; Coimbra; Faro; Gaia Shopping; Guimarães Shopping; Leiria Shopping; Madeira; MAR Shopping; Norte Shopping; Oeiras; Santa Catarina; Setúbal; Viseu; Vasco da Gama; e a do próprio Colombo), e uma loja online (Fnac.pt).



Figura 1 – Mapa das lojas Fnac em Portugal. <http://www.fnacdarty.com>

A Fnac Portugal tem em sua base mais de 500.000 Aderentes – que são os Clientes que possuem o Cartão de Fidelidade Fnac e obtém uma série de vantagens como 5% de desconto em Cartão, 10% de desconto imediato em livros e papelaria, Campanhas exclusivas, entre outras. A adesão pode ser feita também pelo Cartão de Crédito Fnac, que além das vantagens do Fidelidade, disponibiliza opções de reembolso e 1% em Cashback (em tradução livre, “dinheiro de volta”) das compras fora da Fnac.

Em 2016 o departamento de Marketing da Fnac Portugal passou por uma reformulação e foi adotada a visão Omnicanal em toda atmosfera do sector, desde a configuração do organograma e função dos colaboradores, que passaram a atuar de forma integrada, até o desenho de novos processos, que em determinado ponto passam a cruzar e se tornam interdependentes. O Marketing Digital, por exemplo, além de gerenciar as plataformas online, também participa do planejamento e execução das Campanhas offline, e devido à sua natureza mais próxima das Analíticas, passou a influenciar os esforços offline para que estes fossem orientados por dados.

Uma das Campanhas Omnicanal da Fnac são os DNA – Dias e Noites do Aderente ou “Dia do Aderente”. Os DNA são ações promocionais que acontecem três vezes ao ano, tradicionalmente no último fim de semana do mês em que ocorre. São Campanhas exclusivas para Aderentes em que eles podem usufruir de descontos em lista de produtos que contemplam todas as categorias. Em 2016 os

DNA Fnac aconteceram em três ocasiões, 27 a 28 de maio, 26 a 27 de agosto, e 28 a 29 de outubro. Várias plataformas foram implicadas na promoção dos DNA, entre elas o Email Marketing.



Figura 2 – Peça publicitária dos DNA Fnac de outubro de 2016. <http://www.dnoticias.pt>

Apesar da Fnac já segmentar a sua base – atualmente classifica seus Aderentes em 9 Clusters conforme o perfil de compra: Premium; Famílias; Ecléticos; Culturais; Técnicos; Fãs Música; Fnac Kids; Ocasionais; e Sem Segmento, este autor entende ser relevante a análise de grupos de clientes que efetuaram uma transação Omnicanal afim de identificar padrões e formular regras diferentes das aplicadas para a segmentação regular.

Traçar perfis de clientes valiosos que chegam às compras pela abordagem Omnicanal pode ser vital para o Retalho “Click-and-mortar” no contexto da digitalização das transações comerciais. Até esta fase de entrega da Dissertação, a Fnac Portugal e o Retalho de uma forma geral ainda não realiza segmentações e outras análises dentro da abordagem Omnicanal que inclua canais online e offline.

2. RELEVÂNCIA DO TEMA

A estratégia Omnicanal no Retalho é uma evolução da abordagem Multicanal; enquanto a última implica em uma divisão entre o online e o offline, entre o E-commerce e a loja física, entre um canal de promoção e o outro, a segunda permite que o cliente possa mover-se livremente entre os canais – E-commerce, mobile (em tradução livre, “móvel”), ou loja física, dentro do mesmo processo transacional.

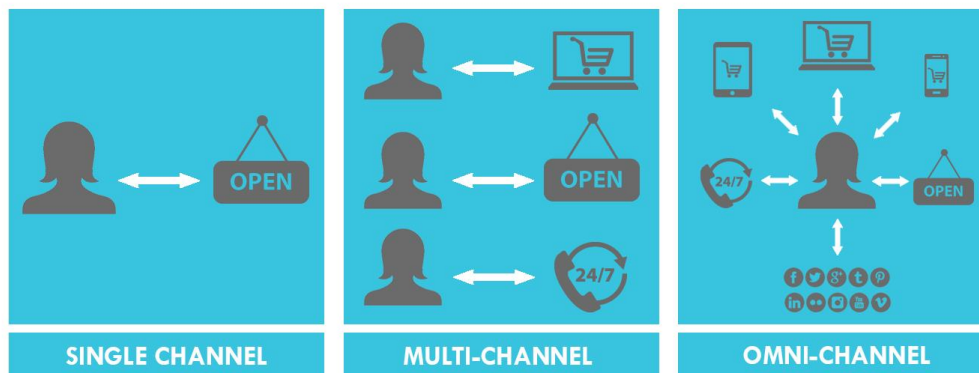


Figura 3 – Omni-channel: Different Channel, Same Message. <http://thedashblog.com>

Para Piotrowicz e Cuthbertson (2014), esta mudança está sendo motivada pela natureza integrativa das novas tecnologias, desde devices (em tradução livre, “dispositivos”) como os smartphones até os softwares de alto nível relacionados ao Omnicanal como por exemplo os aplicativos mobile – “Apps”, as soluções de pagamento mobile, as Redes Sociais, e as soluções de CRM Omnicanal, todas alavancadas pelo Cloud Computing. Todavia, mesmo com esta mudança causada por tecnologias emergentes, que entre outras vantagens permite a otimização de processos a custos reduzidos, a integração dos canais ainda é percebida como um grande desafio pelos Retalhistas.

A falta de consenso quanto ao futuro dos canais digitais, a falta de uma visão unificada da experiência dos Clientes entre os canais, questões relacionadas ao custo envolvido na infraestrutura necessária, adoção de processos ainda não consolidados pela literatura organizacional (mesmo que no campo das TI) e a dificuldade para selecionar a solução ou o conjunto de soluções adequadas às necessidades dos negócios, são apontados como fatores que levam à resistência para aplicação da estratégia Omnicanal no Retalho (PIOTROWICZ; CUTHBERTSON, 2014).

No entanto, conforme discorre Rigby (2011), o Retalho Omnicanal ou Retalho Digital é um caminho sem volta, e as grandes empresas com exceção daquelas que já nasceram “Pure-Digital” como a Amazon, ainda funcionam em sua maioria de forma analógica, e mesmo que tenham suas lojas online, a maioria ainda não soube implementar a estratégia Omnicanal. O autor diz ainda que “Retalhistas Digitais impulsionam a inovação por gastar pesadamente em recrutamento, salários e bônus para atrair e reter talentos técnicos superiores. Eles também estão entre os primeiros a utilizar

o Cloud Computing (que reduz drasticamente os custos de entrada e de operação) e para aumentar a eficiência do marketing através de redes sociais e publicidade online” (RIGBY, 2011).

Ao citar novamente a Amazon, Rigby (2011) indica a importância das inovações para o Retalho Omnicanal, muitas delas pertencente ao campo do Machine Learning e do Data Mining, como os sistemas de personalização e recomendação, além de valiosas patentes sobre inovações essenciais, como o checkout de 1-Click (que encerra a transação em apenas um clique) e um sistema online que permite aos consumidores trocar presentes indesejados antes mesmo de recebê-los.

Em um cenário em que grande parte do Retalho ainda resiste à implementação da estratégia Omnicanal, a criação de um modelo de Clustering que permitirá fazer análises detalhadas dos clientes, apontar suas necessidades e preferências, desenhar segmentos bem definidos, e principalmente, criar diretrizes que permitam otimizar a alocação de recursos entre os canais utilizados nas Campanhas, pode ser um estímulo fundamental para superar esta resistência.

A literatura acadêmica relacionada à estratégia Omnicanal do Retalho ainda é escassa, e as principais referências são publicações ligadas exclusivamente ao Retalho e ao E-commerce, como Journal of Retailing e International Journal of Electronic Commerce, além de publicações com viés acadêmico, mas voltada para os atores dos negócios, como Harvard Business Review e MIT Sloan Management Review. Isto é decorrência de o tema ser relativamente novo e, enquanto método, ter sido pouco testado e atestado.

Este estudo também pode ser de grande valia para profissionais de marketing que trabalham especificamente com CRM ou Database Marketing (em tradução livre, “Marketing de Base de Dados”), não apenas no Retalho, mas em outras indústrias, para a constatação analítica da efetividade do Email Marketing, uma importante ferramenta do Marketing Digital, nas vendas de lojas físicas. De acordo com uma pesquisa feita pelo site Venture Beat (2015) o Email é a ferramenta que apresentou maior ROI – Return on Investment, em tradução livre “Retorno sobre o Investimento” dentre os vários canais de marketing contemplados pelo estudo, como Mídias Sociais, SEM – Search Engine Marketing, TV, rádio, entre outros. Comparado ao Twitter, o Email teve na média uma taxa de Clique de 3%, contra 0.5% da Mídia Social.

3. OBJETIVOS

3.1. OBJETIVO PRINCIPAL

O objetivo principal desta Dissertação é a proposição de um modelo para a segmentação de clientes dentro da abordagem Omnicanal do Retalho. Vários tipos de análise são utilizados no Retalho para a segmentação, sejam de modelagem preditiva ou descritiva. Como exemplos da primeira podem ser citados os modelos de Regressão e Classificação. Já para a segunda, alguns tipos comuns de análise são o Basket Market Analysis, a Análise RFM – Recência, Frequência e Valor Monetário (que também pode ser aplicada em análise preditiva), e a Análise de Clusters ou Clustering, que será empregada neste trabalho, já que o modelo resultará na criação de segmentos ou grupos homogêneos entre si, o que é precisamente o objetivo fundamental do Clustering.

O fato deste estudo possuir os vieses teórico e técnico permitirá que o modelo possa ser revisado por iniciativas acadêmicas e replicado por empresas, especialmente redes Retalhistas de médio e grande porte que tenham adotado ou pretendem adotar a estratégia Omnicanal, e que de preferência já tenham uma estrutura integrada para a gestão dos dados e uma solução de Data Mining, CRM, Customer Intelligence, ou Business Intelligence, que possuam os algoritmos de Clustering.

3.2. OBJETIVOS ESPECÍFICOS

Para que o objetivo principal seja alcançado, o que corresponde à dimensão estratégica da tese, algumas proposições táticas devem ser estabelecidas. Estas proposições, ou objetivos específicos, são os passos necessários e seus respectivos métodos para a construção do modelo de segmentação que será proposto, e correspondem às etapas do processo de Data Mining; são eles:

3.2.1. Recolha dos Dados

O primeiro Objetivo Específico é chegar à correta Recolha dos Dados. Após a definição do problema e a escolha das variáveis que devem fazer parte do modelo, a extração e a consolidação dos Dados de diferentes fontes são as atividades que compõem este Objetivo. No caso deste trabalho, foram relacionados os Dados de vendas dos Aderentes que compraram nos DNA de outubro de 2016, e de interação digital da base de Aderentes, acionada por Email Marketing para a mesma Campanha.

Por serem de fontes distintas, deve-se então cruzar estas duas relações e chegar à tabela síntese “Aderentes que foram acionados por Email Marketing e compraram em lojas físicas durante o período da Campanha”, que servirá como o input (em tradução livre, “entrada”; neste contexto, “dados de entrada”) para o modelo. Em seguida será apresentada a Metadata original – a descrição das variáveis e seus respectivos tipos de dados.

3.2.2. Preparação e Pré-processamento dos Dados

Este Objetivo corresponde à segunda e à terceira etapas de um processo de Data Mining, e partirá da composição do diagrama analítico na solução SAS Enterprise Miner 9.4, que será a plataforma a ser utilizada para o Data Mining. O diagrama seguirá o modelo SEMMA – Sample, Explore, Modify, Model, e Assess, que embora seja mais utilizado para os modelos preditivos, também é aplicável ao Clustering. Em seguida, ocorrerá a Importação dos Dados de input, resultantes do cruzamento na etapa de Recolha. Devem ser confirmados os Roles (em tradução livre, “funções”) e tipos dos dados após a importação.

A Preparação dos Dados passará pela a Análise das estatísticas básicas dos dados de input. É neste ponto que se verificam a presença de valores omissos, dados incorretos e outliers (dados que se encontram fora da região de interesse do espaço de input).

Já o Pré-processamento dos Dados visará à Transformação de variáveis – neste caso utilizar a transformação mais importante, que é a normalização de variáveis, já que os modelos na sua maioria assumem que as variáveis possuem o mesmo peso e importância, e à Redução do espaço de input – pois o excesso de variáveis pode tornar difícil encontrar grupos com características semelhantes, fenómeno conhecido como “maldição da dimensionalidade”.

3.2.3. Segmentação

Esta etapa é constituída da modelagem propriamente dita. A partir de uma nova verificação das estatísticas básicas, deve-se definir quais variáveis serão utilizadas no modelo. A Seleção do algoritmo é o segundo passo – no caso deste trabalho o algoritmo usado será o k-Means, pela facilidade de implementação, pela rapidez e eficiência computacional, e por responder bem com bases de tamanhos diversos. Por fim, é necessário rodar o algoritmo com diferentes números de Clusters, e excluindo ou acrescentando variáveis, até encontrar o melhor resultado.

3.2.4. Avaliação

O último Objetivo Específico a ser alcançado é a Avaliação, onde acontecerá a Validação da Solução onde uma ou mais técnicas são escolhidas para avaliar o modelo, e o Profiling, ou Desenho do perfil cada segmento. Será verificado o grau de importância das variáveis para o modelo, e ocorrerá a análise dos resultados finais. Nesta etapa será confirmada ou descartada a aplicabilidade do modelo desenvolvido na abordagem Omnicanal do Retailho.

4. REVISÃO DA LITERATURA

Embora a contextualização para o tema central desta Dissertação tenha passado pelo Retalho Omnicanal, e o Customer Relationship Management enquanto disciplina e solução tecnológica, a maior parte das consultas bibliográficas serviram para subsidiar a construção do modelo de segmentação dentro da abordagem Omnicanal do Retalho, a partir do levantamento de artigos científicos e publicações com viés técnico e que também explorassem os conceitos principais ligados ao Data Mining – que é a atividade-chave deste trabalho. Ainda assim, tratando-se do Retalho Omnicanal, o artigo base foi o “The future of shopping” (Rigby, 2011), publicado na Harvard Business Review, e o do CRM, “Customer data integration: Reaching a single version of the truth” (DYCHÉ, J. & LEVY, E., 2011).

A maior parte dos autores relacionados para a Revisão da Literatura foram citados em aulas e materiais das disciplinas de Data Mining e Business Intelligence assistidas por este autor na NOVA IMS. O livro “Data mining techniques: for marketing, sales, and customer support” (BERRY & LINOFF, 2004) foi a fonte mais consultada para esta Dissertação, por apresentar os conceitos fundamentais do Data Mining e suas aplicações no Retalho, o que está diretamente ligado ao Objetivo Principal deste trabalho. Pela perspectiva mais operacional, especialmente das etapas do Data Mining que antecedem a modelagem, as principais publicações consultadas foram “Data preparation for data mining” (PYLE, 1999), e “Predictive data mining: a practical guide” (WEISS, S.M. & INDURKHYA, N., 1998).

5. METODOLOGIA

O problema de pesquisa surgiu para este autor após uma série de observações realizadas junto ao Coorientador desta Dissertação, o Gerente de CRM e Marketing Digital da Fnac Portugal, Luis Miguel Santos, sobre os desafios operacionais e estratégicos que o Retalho Europeu e Mundial possui com a digitalização do setor, que vão além da simples instalação de um E-commerce, e que tornam inevitável a adoção da estratégia Omnicanal, como é o caso da Fnac e de outras grandes redes Retalhistas. Entre os desafios observados, um obteve destaque – medir a influência entre os canais online e offline dentro das campanhas de marketing. Esta observação foi corroborada por revisão da literatura.

No caso da Fnac, e de outras redes Retalhistas, o Email Marketing é uma das principais fontes de tráfego e receita para a loja online. Além disso, é um importante canal entre a marca Fnac e os Aderentes do programa de fidelidade. Todavia ainda não há um processo de Business Analytics que permita analisar a influência do Email Marketing nas vendas de Lojas Físicas da Fnac – o que está diretamente relacionado à observação citada acima. Apenas são analisados a performance do Email, com métricas de interação provenientes das Web Analytics, e as conversões em vendas na loja online.

A análise do impacto do Email Marketing nas vendas de Lojas Físicas implicaria na otimização das campanhas em diversas frentes, como na seleção de produtos, no Marketing de conteúdo, e principalmente na segmentação da base de clientes que recebem o Email Marketing, que poderia ser mais assertiva e adequada ao comportamento Omnicanal. Partindo deste fluxo, Email Marketing – Loja Física, é possível criar um modelo de segmentação para a abordagem Omnicanal do Retalho?

Para responder ao problema de pesquisa de forma adequada, o orientador desta Dissertação, Prof. Dr. Rui Gonçalves, e este autor, selecionaram como método a Análise de Clusters, a partir da criação de um projeto de Data Mining, na solução SAS Enterprise Miner, o que permitiria a criação de novos segmentos com regras específicas e já com a premissa de serem Omnicanal. Para este estudo a Fnac Portugal autorizou a recolha de dados de vendas dos Aderentes que compraram na Campanha DNA, nos dias 28 e 29 de outubro de 2016, e dados de interação digital da base de Aderentes, acionada por Email Marketing para esta Campanha.

5.1. RECOLHA DOS DADOS

A extração dos dados foi feita com a solução SAS Guide, ligada a um Data Warehouse Teradata. São registos de Aderentes do programa de fidelidade da Fnac. Por serem de fontes iniciais distintas, os dados de VENDAS e do Email Marketing foram recolhidos em tabelas separadas. As variáveis selecionadas para a tabela dos dados de VENDAS durante a Campanha DNA dos dias 28 e 29 de outubro estão representadas na Metadata na Tabela 1 abaixo:

Tabela 1 - Metadata Original

VARIÁVEIS	CONTEÚDO	TIPO
ID	Número de identificação do Cliente	Nominal
CA_Ano	Montante gasto pelo Cliente no último ano	Interval
CA_DNA	Montante gasto pelo Cliente durante a Campanha DNA outubro	Interval
CA_bilheteira	Montante gasto pelo Cliente em Bilheteira	Interval
CA_burotica	Montante gasto pelo Cliente em Burótica	Interval
CA_casa	Montante gasto pelo Cliente em Artigos para Casa	Interval
CA_discos	Montante gasto pelo Cliente em Discos	Interval
CA_filmes	Montante gasto pelo Cliente em Filmes	Interval
CA_foto	Montante gasto pelo Cliente em Foto	Interval
CA_gaming	Montante gasto pelo Cliente em Gaming	Interval
CA_informatica	Montante gasto pelo Cliente em Informática	Interval
CA_inst_musicais	Montante gasto pelo Cliente em Instrumentos Musicais	Interval
CA_jogos	Montante gasto pelo Cliente em Jogos	Interval
CA_livros	Montante gasto pelo Cliente em Livros	Interval
CA_papelaria	Montante gasto pelo Cliente em Papelaria	Interval
CA_som	Montante gasto pelo Cliente em Som	Interval
CA_telecom	Montante gasto pelo Cliente em Telecomunicações	Interval
CA_tv_video	Montante gasto pelo Cliente em TV/Video	Interval
Cluster	Cluster pré-definido pela Fnac	Nominal
Cod_Loja_DNA	Código da Loja em que foi feita a compra no DNA outubro	Interval
Compras_Ano	Número de compras no último ano	Interval
Compras_DNA	Número de compras realizados no DNA outubro	Interval
Loja_Influencia	Código da Loja de Influência do Cliente	Interval
idade	Idade do Cliente	Interval
sex_cliente	Gênero do Cliente	Nominal
data_venda	Dia do DNA outubro em que foi feita a compra (28 ou 29)	Interval

São 26 variáveis consideradas relevantes, sendo **sex_cliente** e **idade** as sociodemográficas, **ID** a primary key (em tradução livre, “chave-primária”), **data_venda** indica a data em que a compra foi feita na Campanha DNA, **Cluster** contém os segmentos pré-definidos pela Fnac – citados na introdução desta tese, **Cod_Loja_DNA** e **Loja_Influencia** indicam os códigos da loja em que a compra foi feita na Campanha e da loja em que o Aderente mais fez compras – a relação com os números das lojas é demonstrada na Tabela 2. As variáveis antecidas por **CA** contém os montantes gastos por Ano, na Campanha DNA, e por departamento; as antecidas por **Compras** indicam a quantidade de compras no último ano e na Campanha DNA.

Tabela 2 – Códigos das Lojas Fnac

Cod_Loja	LOJA
1	COLOMBO
2	NORTESHOPPING
3	CHIADO
4	CASCAIS
5	STA CATARINA
7	ALMADA
8	GAIA
9	ALGARVE
10	COIMBRA
11	FNAC.PT
12	MADEIRA
13	BRAGA
14	ALFRAGIDE
15	UISEU
16	MAR SHOPPING
17	VASCO GAMA
18	GUIMARAES
19	LEIRIA
20	AEROPORTO
21	AMOREIRAS
22	SETUBAL
23	FARO
24	OEIRAS
25	IST
26	SALDANHA

A seleção de variáveis para a tabela com os dados do Email Marketing deveria seguir as seguintes premissas: indicar que o Aderente interagiu, ou seja, que recebeu e abriu o Email. Não foi considerado se o Aderente efetuou clique, uma métrica recorrente em Web Analytics. Além de identificar a Campanha DNA – já que mais de um tipo de Email foi produzido para a Campanha, a data de interação – que deveria constar como antes ou durante os dias da Campanha, e uma Primary Key que fosse a mesma da tabela de VENDAS, o que permitiria realizar o cruzamento. Segue abaixo na Tabela 3 uma amostra dos registos desta base:

Tabela 3 – Registos da tabela de interação do Email Marketing da Campanha DNA

Primary_key	ID_Fnac_pt	Date_track	Time_track	IdMarketing	Category_Url	Label_Url	Identifier_Url
128369847	196735	28/10/2016	00:00:01	178697835	Dias Aderente	Apple	aa96c50
128369836	710209	27/10/2016	23:59:51	178697835	Dias Aderente	Filmes	aa96c4e
128369834	568479	27/10/2016	23:59:49	178697223	Dias Aderente	Livros	aa974a5

Foram selecionadas 8 variáveis: **Identifier_Url**, **Label_Url**, e **IdMarketing** identificam o conteúdo do Email Marketing, relacionado aos produtos anunciados, **Category_Url** indica a Campanha, **Date_track** e **Time_track** apontam o dia e a hora que o Email foi entregue (lembrando que a relação já possui a condição de conter apenas registos que abriram o Email), o **ID_Fnac_pt** é uma das variáveis de identificação dos Aderentes, mas apenas para as operações da loja online, e **Primary_Key** que corresponde à variável **ID** da tabela de VENDAS.

Após a extração das duas tabelas, foi feito realizado na ferramenta Microsoft Excel o cruzamento entre ambas, até chegar à tabela síntese “Aderentes que foram acionados por Email Marketing e compraram em lojas físicas durante o período da Campanha DNA, nos dias 28 e 29 de outubro de 2016”. Antes das fases de Preparação e Pré-Processamento, que antecedem o modelo e são realizadas já no SAS Enterprise Miner, foram excluídos ainda em Excel os registos de Aderentes que fizeram compras **apenas** na loja online, Fnac.pt, de código 11. Este autor decidiu manter um registo de transação por **ID**, removendo registos de compras duplicados e de compras recorrentes dentro da mesma Campanha (priorizando a compra de maior valor monetário). A relação final, que serão os dados de input do projeto em SAS Enterprise Miner, resultou em **14.792** registos.

5.2. PREPARAÇÃO E PRÉ-PROCESSAMENTO DOS DADOS

Antes da montagem do diagrama analítico no modelo SEMMA – Sample, Explore, Modify, Model, Assess, foi realizada a importação dos dados para servirem como base de input. Como a relação estava em extensão de Excel **.xlsx**, um diagrama paralelo foi criado com os nodes (em tradução livre, “Nós”) **File Import** e **Save Data**. O primeiro operou a importação propriamente dita e o segundo converteu a relação em formato SAS **.sas7bdat** e a disponibilizou em uma SAS Library para que fosse utilizada como input.

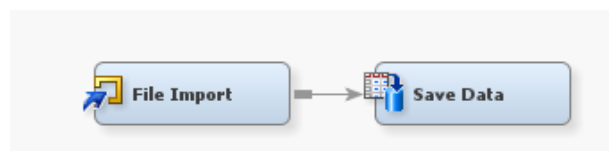


Figura 4 – Nodes File Import e Save Data. SAS Enterprise Miner 9.4

4.2.1. Sample

Já como input do projeto, foram definidos os Roles e Levels (em tradução livre, “Níveis”; neste contexto, “Tipos”) das variáveis, que são as funções e os tipos de dados. Todas as variáveis estavam já definidas com o Role Input, com exceção da ID e data_venda, para as quais foram corretamente atribuídos os Roles de ID e Time ID respetivamente. Variáveis nominais e intervalares foram corretamente classificadas pelo node **Input**, não havendo necessidade de alterações nesta etapa.

Tabela 4 - Metadata com os Roles e Levels

Variable Name	Role	Measurement Level
CA_Ano	Input	Interval
CA_DNA	Input	Interval
CA_bilheteira	Input	Interval
CA_burotica	Input	Interval
CA_casa	Input	Interval
CA_discos	Input	Interval
CA_filmes	Input	Interval
CA_foto	Input	Interval
CA_gaming	Input	Interval
CA_informatica	Input	Interval
CA_inst_musicais	Input	Interval
CA_jogos	Input	Interval
CA_livros	Input	Interval
CA_papelaria	Input	Interval
CA_som	Input	Interval
CA_telecom	Input	Interval
CA_tv_video	Input	Interval
Cluster	Input	Nominal
Cod_Loja_DNA	Input	Interval
Compras_Ano	Input	Interval
Compras_DNA	Input	Interval
ID	ID	Nominal
Loja_Influencia	Input	Interval
data_venda	Time ID	Interval
idade	Input	Interval
sex_cliente	Input	Nominal

4.2.2. Explore

Ao analisar estatísticas básicas das variáveis nominais e intervalares com o node **StatExplore**, este autor pôde assimilar informações relevantes sobre a composição da base que poderiam guiar a redução do espaço de input e fornecer insights (em tradução livre, “discernimentos”) para a escolha de variáveis na fase de modelagem. Segue abaixo as tabelas com as estatísticas:

Tabela 5 – Estatísticas das variáveis de Classe (A)

Data	Variable	Role	Number of Levels	Missing	Mode	Mode Percentage	Mode2	Mode2 Percentage
TRAIN	Cluster	INPUT	9	0	Famílias	20.93	Culturais	19.22
TRAIN	sex_cliente	INPUT	3	36	M	54.73	F	45.03

Tabela 6 – Estatísticas das variáveis de Classe (B)

Data Role	Variable Name	Role	Level	Percent	Frequency Count
TRAIN	Cluster	INPUT	Famílias	20.93023	3096
TRAIN	Cluster	INPUT	Culturais	19.21985	2843
TRAIN	Cluster	INPUT	Fãs Música	13.06111	1932
TRAIN	Cluster	INPUT	Técnicos	11.56706	1711
TRAIN	Cluster	INPUT	Premium	11.16144	1651
TRAIN	Cluster	INPUT	Fnac Kids	8.998107	1331
TRAIN	Cluster	INPUT	Ocasionais	8.896701	1316
TRAIN	Cluster	INPUT	Ecléticos	5.110871	756
TRAIN	Cluster	INPUT	Sem segmento	1.054624	156
TRAIN	sex_cliente	INPUT	M	54.72553	8095
TRAIN	sex_cliente	INPUT	F	45.0311	6661
TRAIN	sex_cliente	INPUT		0.243375	36

Observa-se que a maior parte dos Aderentes faz parte dos Clusters pré-definidos pela Fnac, **Famílias**, (20,93%), e **Culturais**, (19,22%). Ainda é cedo para adiantar causalidades e explorar o fenómeno Omnicanal, mas a experiência deste autor em análises do Retalho indica que estas fatias provavelmente serão semelhantes se a distribuição for aplicada à base total de Aderentes, devido à liderança da Fnac em vendas de Livros, e à experiência familiar de compra nas lojas físicas. Há um equilíbrio notável entre **Homens** (54,72%) e **Mulheres** (45%). Os missing values (em tradução livre, “valores omissos”) presentes na variável **sex_cliente** são irrelevantes para esta análise.

Tabela 7 – Estatísticas das variáveis Intervalares

Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
CA_Ano	INPUT	847.3166	4649.128	14792	0	0.894	485.2559	487829.9	88.34493	8704.446
CA_DNA	INPUT	115.7698	638.0831	14792	0	0.041	26	49467.99	63.09472	4793.685
CA_bilheteira	INPUT	61.66938	143.9742	14792	0	0	0	4892.2	7.314896	133.2156
CA_burotica	INPUT	2.279563	18.00188	14792	0	0	0	406.496	9.542702	106.6877
CA_casa	INPUT	4.519232	35.62838	14792	0	0	0	1002.406	12.93777	212.2672
CA_discos	INPUT	45.08503	217.8925	14792	0	0	0	9706.276	17.49218	481.5651
CA_filmes	INPUT	27.7788	112.832	14792	0	0	0	2867.211	11.48159	197.0529
CA_foto	INPUT	42.6236	178.8438	14792	0	0	0	6687.415	11.1726	230.5222
CA_gaming	INPUT	45.48493	188.5769	14792	0	0	0	15700.15	42.12605	3263.302
CA_informatica	INPUT	262.2937	3464.798	14792	0	0	16.252	343118.9	85.68423	7819.343
CA_inst_musicais	INPUT	4.225949	40.19364	14792	0	0	0	3194.156	40.46601	2786.015
CA_jogos	INPUT	14.60962	38.70005	14792	0	0	0	1259.584	7.122936	112.8359
CA_livros	INPUT	129.4375	216.9368	14792	0	0	64.443	4192.683	6.159574	66.72377
CA_papelaria	INPUT	9.231897	23.50308	14792	0	0	0	726.3625	7.95658	121.3426
CA_som	INPUT	25.5317	99.19582	14792	0	0	0	3983.732	14.63838	386.8261
CA_telecom	INPUT	122.8205	407.2882	14792	0	0	0	20023.23	19.09337	683.9331
CA_tv_video	INPUT	46.30587	1091.204	14792	0	0	0	129649.8	113.6112	13460.33
Cod_Loja_DNA	INPUT	10.05692	7.19912	14792	0	1	9	26	0.358281	-1.0714
Compras_Ano	INPUT	15.01075	16.4149	14792	0	1	11	323	4.29655	38.03769
Compras_DNA	INPUT	1.206598	0.561819	14792	0	1	1	10	4.09139	26.57027
Loja_Influencia	INPUT	10.08964	7.03208	14659	133	1	10	26	0.342111	-1.02032
idade	INPUT	43.99452	11.87163	14782	10	18	44	90	0.308054	-0.14278

A análise da tabela de estatísticas das variáveis intervalares mostra que a idade média dos Aderentes é de **44 anos**, sendo a idade mínima 18 e a máxima 90 anos. Esta observação reforça o possível carácter familiar da experiência de compra na Fnac, especialmente nos DNA – tendo em vista que além da abordagem promocional, a Campanha aconteceu em um fim de semana. Verifica-se que a variável **Loja_Influencia** possui 133 missing values (em tradução livre, “valores omissos”) que serão tratados posteriormente.

Na variável **Compras_Ano** o valor máximo é de 323, o que indica um possível outlier, pois a média anual é de 11 compras. Porém a observação mais alarmante está nos montantes gastos por departamento – só há informação estatisticamente relevante para as categorias Livros e Informática. A diferença com relação às outras variáveis é tão grande que o agrupamento das demais seria insignificante para a análise. Partindo das premissas que os montantes foram gastos no último ano, e que não foi constatado erro na revisão da fase de recolha, diversas hipóteses podem ser levantadas, e a resolução deste problema desviaria do objetivo desta Dissertação.

O autor e o Coorientador decidiram então retornar à etapa de Preparo e Pré-processamento dos dados, e além de excluir as variáveis dos montantes gastos por departamento, implementaram

outras medidas para reduzir ao máximo o espaço de input, ainda antes da importação para o SAS Enterprise Miner, e assim simplificar o processo de modelagem.

5.3. NOVA PREPARAÇÃO E PRÉ-PROCESSAMENTO DOS DADOS

Além da exclusão das variáveis dos montantes gastos por departamento, também foram descartadas as variáveis **ID** – já que os dados de output (em tradução livre, “dados de saída”) neste caso não serão usados pela Fnac para fins de Campanha, **Cluster**, e **Compras_DNA**. Foi decidido também transformar as variáveis em nominais (o algoritmo irá posteriormente transformar os dados em numéricos na fase de modelagem) com exceção da variável **data_venda**, que apenas teve a substituição dos valores 28/10/2016 e 29/10/2016 por **1** e **2**, respetivamente. A variável **ex-cliente** permaneceu com a mesma configuração, mas teve os 36 registos com missing values excluídos. As demais variáveis foram divididas por classes após o desenho da distribuição de cada uma, conforme os gráficos abaixo:

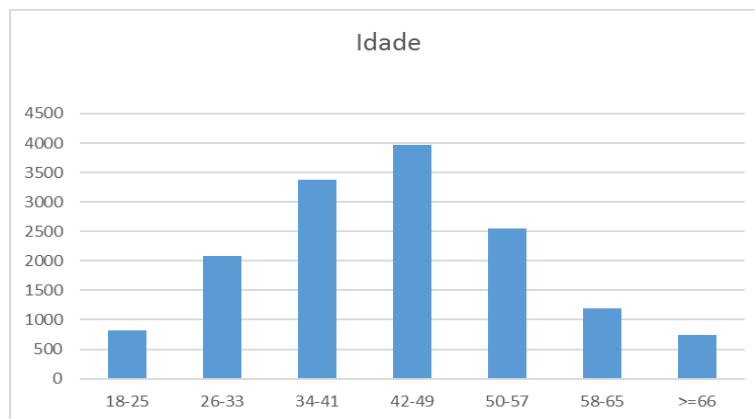


Gráfico 1 – Distribuição da variável idade

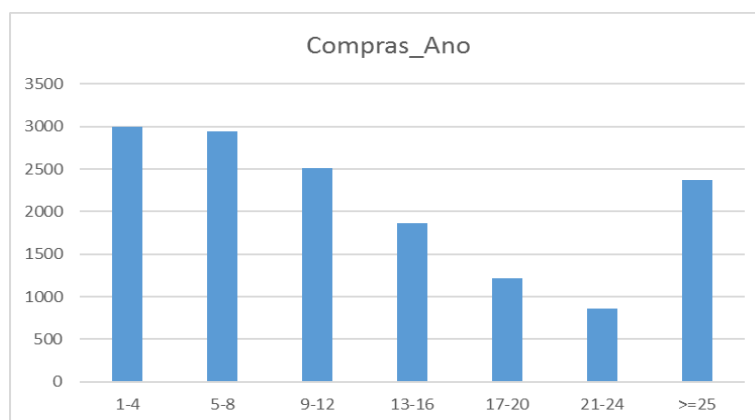


Gráfico 2 – Distribuição da variável Compras_Ano

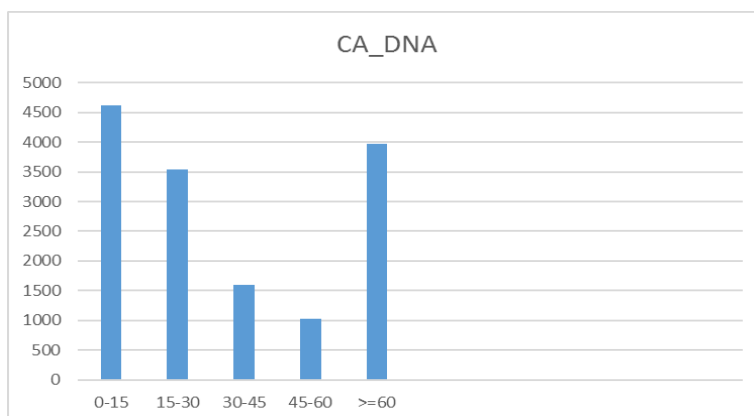


Gráfico 3 – Distribuição da variável CA_DNA

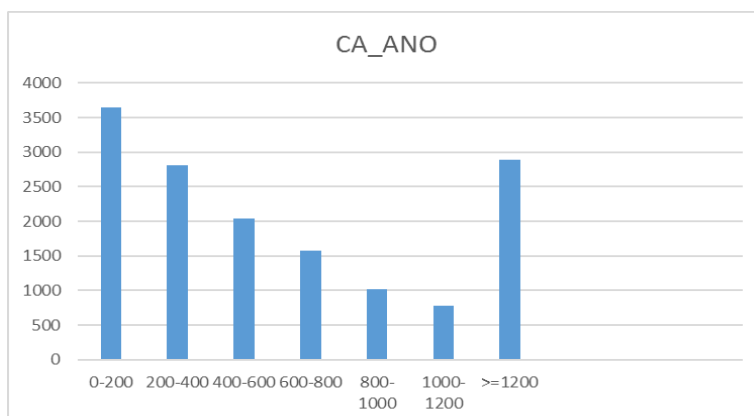


Gráfico 4 – Distribuição da variável CA_ANO

As variáveis **Cod_Loja_DNA** e **Cod_Loja_Influencia** tiveram seus valores agrupados pela região da localização das lojas de acordo com a Nomenclatura de Unidades Territoriais para Fins Estatísticos de Nível II (NUTS II), acrescentados de Cascais e Setúbal, e passaram a ter a label (em tradução livre, “rótulo”) **REGIAO_Loja_DNA** e **REGIAO_Loja_Influ**:

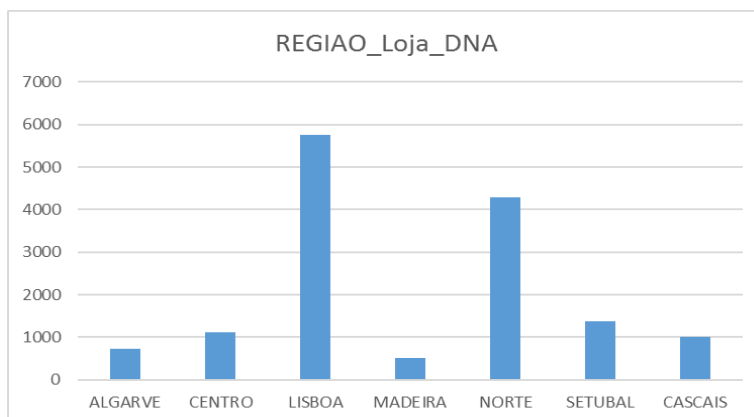


Gráfico 5 – Distribuição da variável REGIAO_Loja_DNA

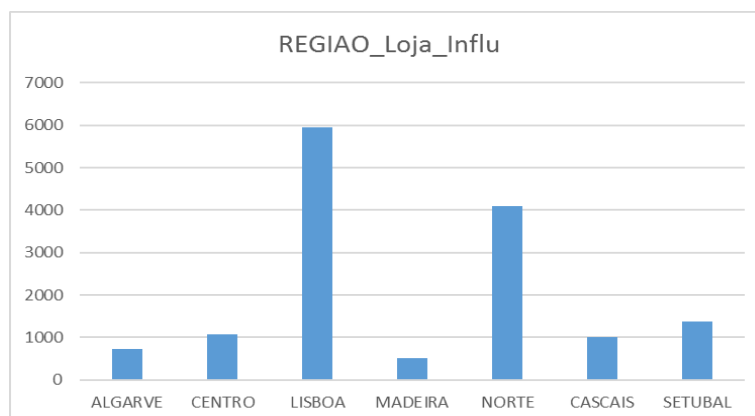


Gráfico 6 – Distribuição da variável REGIAO_Loja_Influ

4.3.1. Sample

Foi criado um novo projeto em SAS Enterprise Miner e a base importada e convertida em formato SAS. **sas7bdat** para fornecer os dados de input passou a possuir 8 variáveis e 14.756 registos. Os Roles e Levels foram definidos corretamente pelo node **Input**. Segue abaixo a tabela da nova Metadata:

Tabela 8 – Nova Metadata com os Roles e Levels

Variable Name	Role	Measurement Level
Faixa_Compras_Ano	Input	Nominal
Faixa_Gasto_Ano	Input	Nominal
Faixa_Gasto_DNA	Input	Nominal
Faixa_Idade	Input	Nominal
REGIAO_Loja_DNA	Input	Nominal
REGIAO_Loja_Influ	Input	Nominal
data_venda	Input	Interval
sex_cliente	Input	Nominal

4.3.2. Explore

A etapa Explore do processo de Data Mining foi realizada com os nodes **Multiplot** e **StatExplore**. O primeiro permite a análise por gráficos das variáveis e apresentou as mesmas distribuições apresentadas nesta Dissertação para a divisão das variáveis em classes. O segundo apresentou as estatísticas básicas em forma de tabela, como já foi demonstrado na primeira versão do Preparo e Pré-processamento dos dados. Segue a tabela com algumas estatísticas:

Tabela 9 – Estatísticas das variáveis de Classe

Data Role	Variable Name	Role	Number of Levels	Missing	Mode	Mode Percentage	Mode2	Mode2 Percentage
TRAIN	Faixa_Compras_Ano	INPUT	7	0	1 a 4	20.32	5 a 8	19.92
TRAIN	Faixa_Gasto_Ano	INPUT	7	0	0-200	24.70	>=1200	19.59
TRAIN	Faixa_Gasto_DNA	INPUT	5	0	0-15	31.29	>=60	26.92
TRAIN	Faixa_Idade	INPUT	7	0	42-49	26.91	34-41	22.93
TRAIN	REGIAO_Loja_DNA	INPUT	7	0	LISBOA	38.97	NORTE	29.01
TRAIN	REGIAO_Loja_Influ	INPUT	7	0	LISBOA	40.36	NORTE	27.80
TRAIN	sex_cliente	INPUT	2	0	M	54.86	F	45.14

De certa forma estas estatísticas já foram adiantadas pela a análise das distribuições, mas alguns pontos devem ser ressaltados: **20,32%** dos Aderentes realizaram **de 1 a 4 compras** no último ano, seguido de uma fatia com percentagem quase idêntica, **19,92%**, que realizaram **de 5 a 8 compras**. São valores de engajamento normais para o Retalho, mas podem sugerir um comportamento associado às grandes Campanhas, como os DNA, que acontecem 3 vezes ao ano. Observa-se um padrão nas distribuições **Faixa_Compras_Ano**, **Faixa_Gasto_Ano**, e **Faixa_Gasto_DNA**: a frequência diminui gradativamente até surgir um pico na última classe. Esta disposição do histograma sugere a presença de outliers ou a inclusão de dados de outra distribuição. Este autor e o Coorientador optaram por não tratar outliers devido à natureza desta Dissertação.

Outra importante observação é a diferença, em montante gasto no último ano, entre a maior fatia, com gasto de até **200 Euros** no último ano (**24,70%**), e a segunda maior, com gasto maior ou igual a **1200 Euros (19,59%)**, seguindo o padrão citado acima, embora as classes desta variável tenham quase a mesma frequência, o que não acontece com as variáveis **REGIAO_Loja_DNA** e **REGIAO_Loja_Influ**, onde **LISBOA** e **NORTE** respondem por mais de **70%** em ambas.

4.3.3. Modify

As transformações de variáveis antecedem a fase de modelação e são fundamentais para o correto processamento dos algoritmos. Uma das principais atividades desta fase é a normalização, que faz com que as variáveis intervalares possuam o mesmo peso, e algoritmos como o K-Means, utiliza a distância Euclidiana e depende de as variáveis estarem em escala idêntica.

As demais atividades de transformação foram realizadas antes da importação dos dados de input, conforme já foi descrito. Não houve disponibilidade de variáveis que permitissem a criação de outras, com lógicas de negócios ou com finalidades estatísticas, todavia para chegar ao objetivo desta Dissertação não será necessária a introdução de novas variáveis.

5.4. SEGMENTAÇÃO

4.4.1. Model

A parte mais importante e que requer mais acurácia por parte do investigador, ou analista, é a modelagem. Tratando-se da proposição de um modelo de segmentação, que terá caráter apenas exploratório, não é necessária a comparação entre os vários algoritmos possíveis para o Clustering. Conforme já foi adiantado, o algoritmo que será usado no modelo é o K-Means.

De acordo com Sofia, H. & Henriques, R. (2013) o K-Means é um dos mais usados algoritmos entre os “data miners” e isso se dá, dentre outros fatores, por sua facilidade de implementação e a capacidade de lidar com bases de dados de tamanhos diversos. Este algoritmo atua por iteração, para definir os centroides dos Clusters, também chamados seeds (em tradução livre “sementes”), que são os pontos de características médias dos mesmos.

O processo de inicialização para a busca do centroide pelo algoritmo K-Means é feita de forma aleatória. Estando definidos os centroides, os objetos restantes são agrupados ao Cluster cujo centroide se encontra mais próximo, ou seja, ao que mais se assemelham. Após estes agrupamentos iniciais, é calculada uma média de cada cluster, processo que continua até que o algoritmo estabelece **k** Clusters que tenham máxima semelhança intra-cluster e máxima dissemelhança inter-cluster.

O problema de negócios deve determinar quais variáveis devem ser escolhidas, mas como este trabalho não visa resolver problemas mercadológicos específicos, inicialmente as 8 variáveis serão utilizadas. Como o K-Means necessita que o número de Clusters seja definido previamente, é preciso fazer testes consecutivos, aumentando o número de Clusters e analisando os resultados. Vale ressaltar que o algoritmo gera resultados diferentes cada vez que é rodado, sendo necessária a utilização de um node **Cluster** para cada teste. Foram realizados 6 testes até a verificação do melhor resultado:

- I) Com 3 Clusters – O “Cluster “1” obteve uma dimensão de 58,3%, o “Cluster 2” de 22,5%, e o “Cluster 3” de 19,2%. Apenas duas variáveis foram usadas para as regras de partição, **data_venda** e **sex_cliente**;
- II) Com 4 Clusters – Os Clusters obtiveram dimensões semelhantes: o “Cluster 1” com 22,5%, o “Cluster 2” com 32,4%, o “Cluster 3” com 19,2%, e o “Cluster 4” com 25,9%”. Porém **data_venda** e **sex_cliente** ainda continuam como únicas variáveis usadas para regras de partição;
- III) Com 5 Clusters – A variável **REGIAO_Loja_DNA** passou também a ser usada na partição; novamente os Clusters obtiveram dimensões semelhantes com exceção do “Cluster 3”, que ficou com 3,30% e é composta por Aderentes que tem como loja de influência **MADEIRA** e compraram nos DNA na **mesma loja**;
- IV) Com 4 Clusters e sem a variável **data_venda** – A variável **REGIAO_Loja_Influ** passou também a ser usada na partição; o “Cluster 1” ficou com 15,3% e composto apenas por **mulheres**, o “Cluster 2” com 10,5%, o “Cluster 3” com 27,6% e composto apenas

por **homens**, e o “Cluster 4” com 46,6%; as variáveis **Faixa_Compras_Ano**, **Faixa_Gasto_Ano**, e **Faixa_Idade** tiveram importância nula;

- V) Com 4 Clusters e sem a variável **sex_cliente** – o “Cluster 1” com 4,3% composta apenas por Aderentes com loja de influência e loja de compra nos DNA **ALGARVE**, o “Cluster 2” com 39,7%, o “Cluster 3” com 10,2%, e o “Cluster 4” com 45,8%”. A única variável além de **sex_cliente** a não ser utilizada na partição foi **Faixa_Compras_Ano**, que novamente teve importância nula;
- VI) Com 4 Clusters e sem as variáveis **data_venda** e **sex_cliente** – Foram criados apenas 3 Clusters, apesar da definição de 4 Clusters como número máximo; o “Cluster 1” com 24,3%, o “Cluster 2” com 14,8%, o “Cluster 3” com 60,9%; além das variáveis excluídas, **Faixa_Compras_Ano** e **Faixa_Gasto_Ano** tiveram importância nula e não foram usadas na partição;

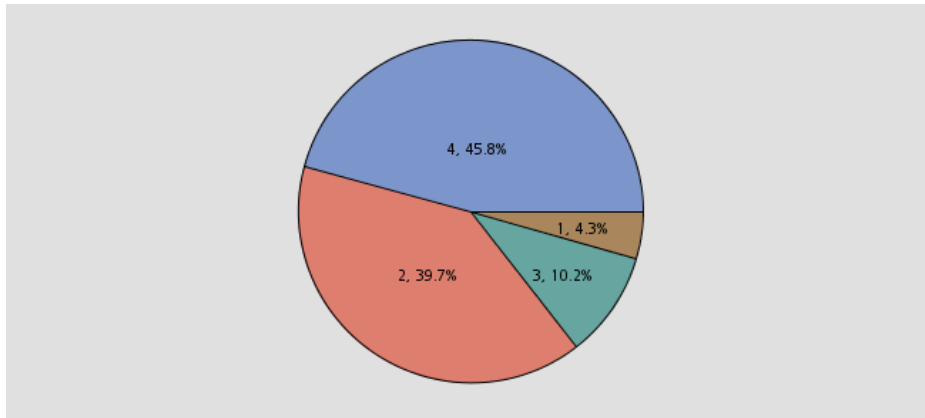


Gráfico 7 – Dimensões dos Clusters. SAS Enterprise Miner 9.4

6. RESULTADOS E DISCUSSÃO

6.1. AVALIAÇÃO

O teste V “Com 4 Clusters e sem a variável **sex_cliente**” obteve o melhor resultado pois apresentou o maior número de variáveis utilizadas nas regras de partição, o que trará maior refinamento para a definição das regras dos clusters, e apenas duas variáveis com importância nula. Segue abaixo a representação em Árvore de Decisão da formação dos Clusters, e o gráfico circular com as dimensões dos mesmos:

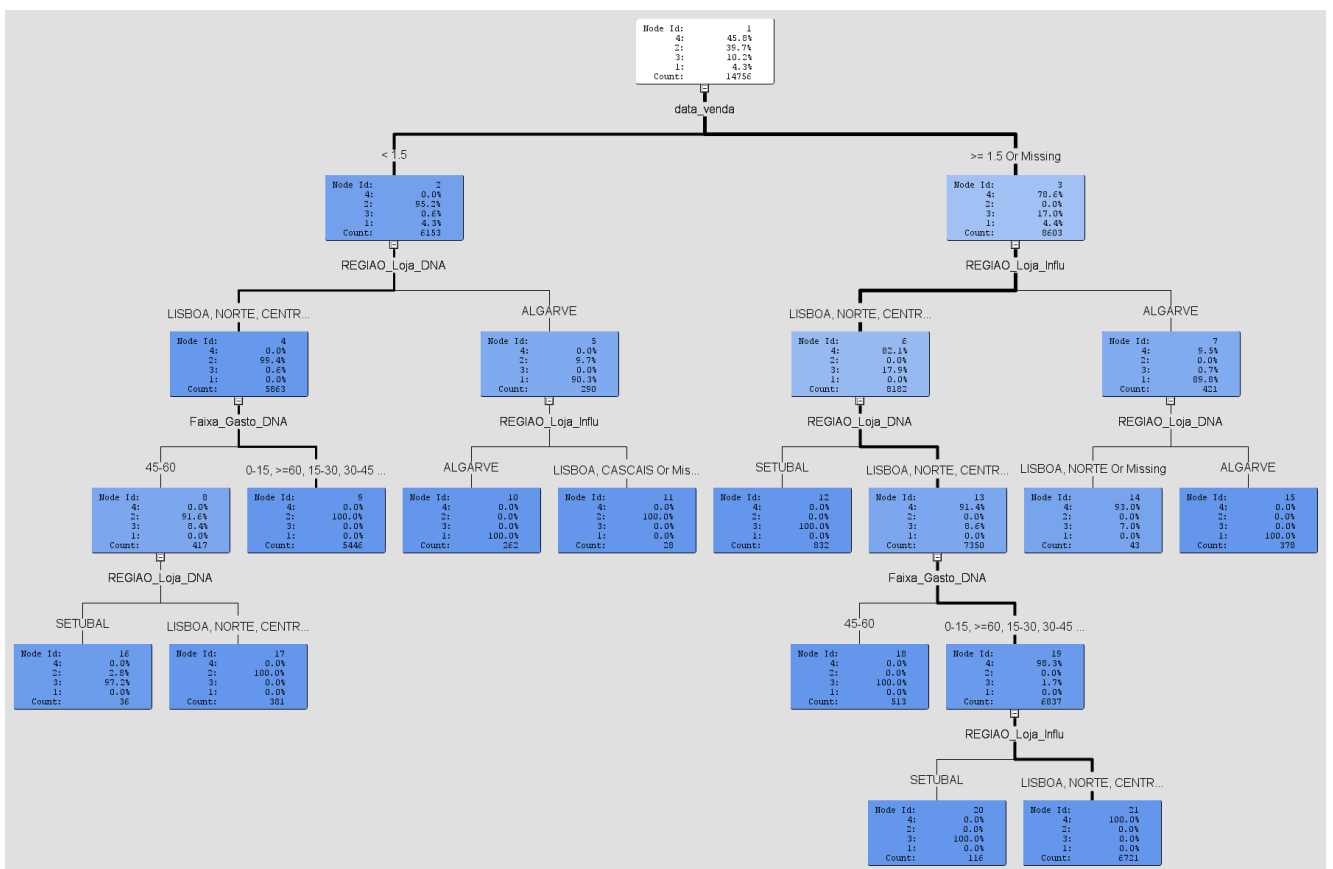


Figura 5 – 4 Clusters sem a variável **sex_cliente**. SAS Enterprise Miner 9.4

Esta etapa corresponde à fase Assess do processo SEMMA. A técnica escolhida por este autor para iniciar a avaliação do modelo e a interpretação dos Clusters é a Multidimensional Scaling (Figura 4), que consiste na análise das distâncias entre os Clusters dispostos em mapa proveniente da Matriz de distância (Tabela 10). Observa-se uma grande distância do Cluster 1 para os demais, o que antecipa sua maior peculiaridade. Os Clusters 2, 3, e 4, embora mais próximos, preservam uma distância que garante a heterogeneidade entre eles. A Multidimensional Scaling está a sugerir que o modelo foi desenvolvido corretamente. O próximo passo da Análise é o Profiling, em tradução livre “definição do perfil”, de cada Cluster.

Tabela 10 – Matriz de distância entre os Clusters

SEGMENT	_1	_2	_3	_4
1	0	1.705376	1.757861	1.678293
2	1.705376	0	1.310717	1.014679
3	1.757861	1.310717	0	1.033504
4	1.678293	1.014679	1.033504	0

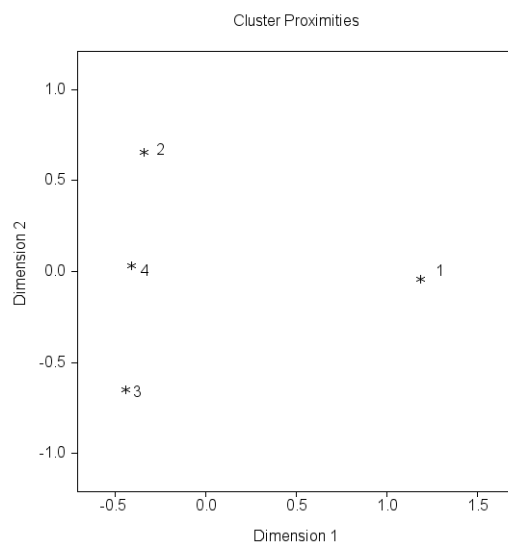


Figura 6 – Mapa obtido por Multidimensional Scaling. SAS Enterprise Miner 9.4

6.2. PROFILING

Cluster 1 – Possui a menor dimensão entre os 4, representando 4,3% do total de Aderentes. Conforme foi verificado com a Multidimensional Scaling, tende a ser muito diferente dos demais, a começar pela variável, neste caso variáveis, que mais os discrimina: **REGIAO_Loja_Influ** e **REGIAO_Loja_DNA**. Abaixo está o gráfico de Variable Worth do Cluster 1:

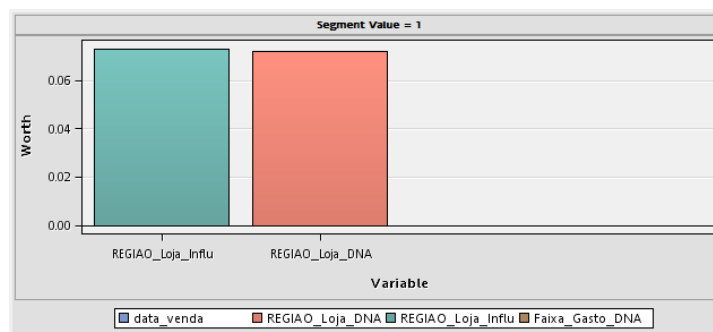


Gráfico 8 – Variable Worth do Cluster 1. SAS Enterprise Miner 9.4

Conforme já foi apontado na fase de modelagem, este Cluster é composto por Aderentes com loja de influência e que compraram na Campanha DNA de outubro de 2016 nas lojas do ALGARVE – Fnac Albufeira e/ou Fnac Faro.

Cluster 2 – É o segundo maior em dimensão, com 39,7% da base total. Possui como variável que mais o discrimina a **data_venda**, conforme os Gráficos 9 e 10 abaixo. Todos os Aderentes deste Cluster compraram nas lojas físicas no primeiro dia dos DNA, 28/10/2016, sexta-feira; 42% dos Aderentes compraram na região de Lisboa e 31% nas lojas do Norte. A mesma tendência segue para as lojas de influência.

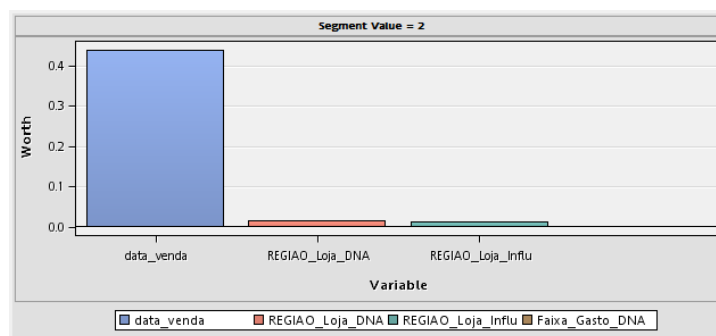


Gráfico 9 – Variable Worth do Cluster 2. SAS Enterprise Miner 9.4

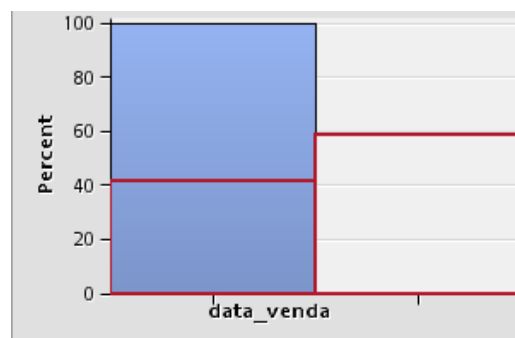


Gráfico 10 – Histograma da variável **data_venda** do Cluster 2. SAS Enterprise Miner 9.4

Cluster 3 – Apresenta uma dimensão de 10,2% da base total; as variáveis mais discriminantes deste Cluster são **REGIAO_Loja_DNA**, **REGIAO_Loja_Influ**, **Faixa_Gasto_DNA** e **data_venda**. 58% dos Aderentes deste Cluster compraram em SETÚBAL, e 23% em LISBOA; 56% tem a loja de influência também em SETÚBAL. 40% gastaram entre 45 e 60 Euros nos DNA, e esta foi a classe de menor frequência na base total. Quase todos os Aderentes deste Cluster compraram no segundo dia dos DNA, 29/10/2016, sábado.

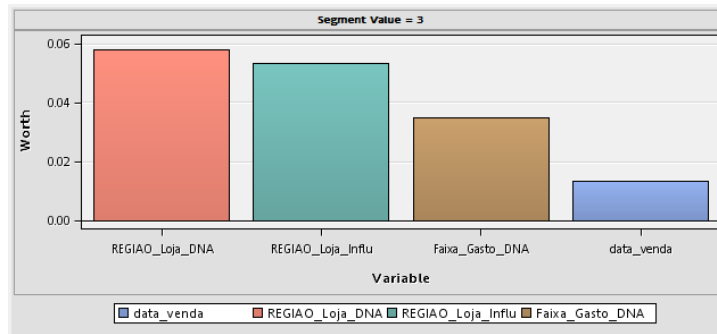


Gráfico 11 – Variable Worth do Cluster 3. SAS Enterprise Miner 9.4

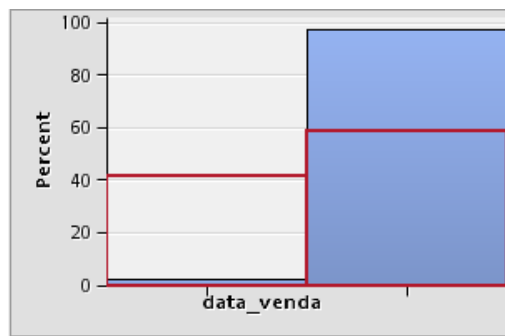


Gráfico 12 – Histograma da variável **data_venda** do Cluster 3. SAS Enterprise Miner 9.4

Cluster 4 – Possui a maior dimensão, com 45,8% do total da base de Aderentes. A variável que mais discrimina é a **data_venda**, todos os Aderentes do Cluster 4 compraram no dia 29/10/2016, sábado. 44% dos Aderentes compraram na região de Lisboa e 34% nas lojas do Norte. A mesma tendência segue para as lojas de influência. Ao contrário do Cluster 3, não foram registadas compras entre 45 e 60 Euros para o Cluster 4. A faixa de gastos de maior frequência foi “até 15 Euros”, 34%.

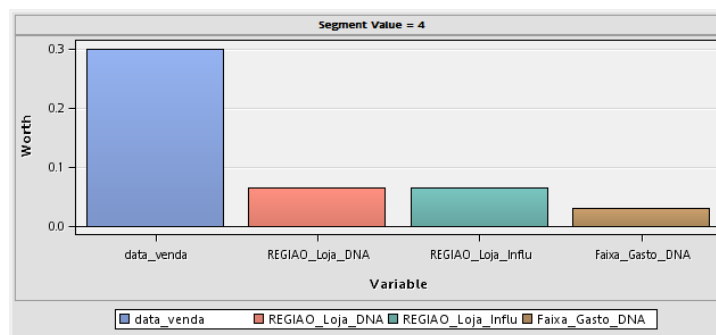


Gráfico 13 – Variable Worth do Cluster 3. SAS Enterprise Miner 9.4

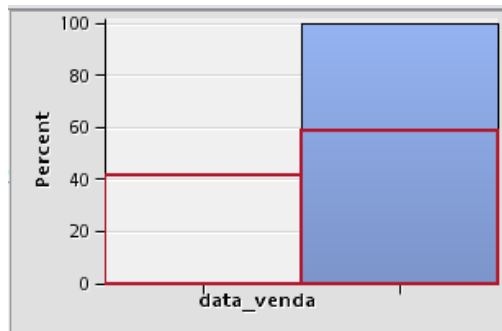


Gráfico 14 – Histograma da variável **data_venda** do Cluster 4. SAS Enterprise Miner 9.4

A análise prévia da Multidimensional Scaling foi confirmada pelo Profiling dos Clusters. O Cluster 1 é de fato o de maior peculiaridade, e mostra que Aderentes Fnac residentes no ALGARVE, entre as regiões que não sejam LISBOA e NORTE, apresentaram a maior tendência para o comportamento de compra Omnicanal, justificando uma investigação mais apurada do fenómeno e a otimização dos esforços para este segmento de Aderentes em campanhas que envolvam canais digitais e não-digitais. Os Clusters 2 e 4 são os mais semelhantes entre si e em relação às estatísticas da base total, mas mostraram claramente que estão mais propensos para a conversão de acordo com o dia.

O Cluster 3 embora tenha acompanhado parte das tendências do 2 e do 4, apresentou um número considerável de Aderentes Fnac na região de SETÚBAL e uma frequência alta na faixa de gasto entre 45 e 60 Euros, o que sugere que Aderentes valiosos e que possuem o comportamento Omnicanal estão inseridos neste Cluster.

7. CONCLUSÕES

O modelo para a segmentação dentro da abordagem Omnicanal do Retalho, desenhado por este autor e corroborado por fundamentos teóricos provenientes da revisão bibliográfica, apresentou o comportamento esperado e demonstrou ser aplicável aos esforços de marketing de empresas retalhistas, guardadas as condições de infraestrutura e expertise analítica necessárias. Ele foi testado seguindo os processos de um projeto padrão de Data Mining, e mesmo que a fase de modelagem tenha sido feita através do processo SEMMA, desenvolvido pela SAS especificamente para o Enterprise Miner, poderá ser aplicado em outras plataformas como Weka e Rapid Miner por exemplo, que possuem os mesmos algoritmos para o Clustering, por métodos de otimização ou hierárquicos, podendo apenas haver limites quanto à dimensão das bases e às ferramentas analíticas disponíveis.

A Recolha dos Dados foi realizada conforme o planejado; a consolidação dos dados provenientes de fontes distintas foi facilitada pela infraestrutura de gestão de dados da Fnac Portugal, principalmente pelo uso da plataforma SAS Guide, que além de permitir a extração dos dados de forma intuitiva e orientada para projetos – não havendo a necessidade de usar as linguagens SAS ou SQL, permite uma série de análises estatísticas que garantem a qualidade dos dados pré-extração. Embora o Guide e o Enterprise Miner façam parte do mesmo conjunto de soluções SAS, no caso deste trabalho as plataformas não estavam interligadas – as operações com o Guide foram feitas na sede da Fnac Portugal, e o desenvolvimento do projeto de Data Mining com o Enterprise Miner nos laboratórios da NOVA IMS e de forma remota por este autor.

A seleção das variáveis para o input, demonstradas na apresentação da Metadata Original, foi baseada nos conjuntos de dados utilizados no Retalho para a segmentação, e contemplou dados sociodemográficos, como género, idade e localização das lojas, e dados de históricos de compras, como montante gastos por ano, na Campanha, e por departamento; além de uma variável que indicava um Cluster pré-definido pela Fnac, o que poderia contribuir para insights durante a construção do modelo.

Foram escolhidas 26 variáveis, e embora esta seja uma quantidade excessiva, levantada principalmente pelo número de departamentos da Fnac, havia uma expectativa de que diversas variáveis seriam cortadas para a redução do espaço de input quando fossem aplicadas matrizes de correlação. Porém, já na análise das estatísticas básicas foi constatada uma alta discrepância entre os montantes gastos nos departamentos Livros e Informática, e os demais. De tal forma que o simples agrupamento dos departamentos restantes não seria suficiente para que a análise fosse relevante.

As variáveis dos montantes gastos por departamento foram então cortadas, e para reduzir ao máximo o espaço de input, outras variáveis também o foram, como **ID** e **Cluster**. Das 8 variáveis restantes, com exceção da **sex_cliente** e da **data_venda**, que indicam género e data da compra na Campanha respetivamente, todas as outras foram divididas em classes a partir do desenho da distribuição de cada uma. Esta é uma situação inesperada para o analista, um “um cisne negro” no processo de Data Mining, mas sobre a qual ele pode ter controle utilizando técnicas estatísticas como a citada acima, que não resultem em perda da qualidade do modelo. Caso não seja possível prosseguir, deve-se retornar ao processo de recolha e começar novamente.

O algoritmo K-Means já havia sido definido previamente por este autor para o Clustering devido à sua facilidade de implementação, rapidez e eficiência computacional, e por responder bem com bases de tamanhos diversos, que são características dos algoritmos por Método de Partição ou Otimização. Estas qualidades foram confirmadas no desenvolvimento do modelo: a normalização, para que os dados tenham o mesmo peso e importância, a redução do espaço de input, e a definição prévia do número de Clusters, pré-requisitos para rodar o K-Means, foram tarefas executadas sem muitas dificuldades. O baixo custo computacional pôde ser notado na rapidez de processamento entre as execuções dos testes, e o tamanho da base não foi um empecilho para a performance do algoritmo.

Foram realizados 6 testes até a constatação do melhor resultado, que teve a configuração de “4 Clusters e sem a variável **sex_cliente**”, e que apresentou o maior número de variáveis utilizadas nas regras de partição. A técnica escolhida para a validação do modelo e interpretação dos Clusters foi a Multidimensional Scaling – em que um mapa é gerado a partir da matriz de distância entre os Clusters, sendo possível analisar visualmente as dissimilaridades entre os mesmos. Outras técnicas poderiam ter sido utilizadas com a utilização de gráficos de linha e de dispersão, mas a Multidimensional Scaling se mostrou eficaz ao ser corroborada pela definição do perfil dos Aderentes dos Clusters, como por exemplo ao verificarmos no mapa que o Cluster 1 é o mais distante entre os demais, o que indica sua heterogeneidade inter-cluster – o que foi comprovado com a análise posterior.

A aplicabilidade do modelo foi atestada pelo alcance dos Objetivos Específicos, mesmo com o surgimento de fatores inesperados, como a necessidade de reformulação da etapa de Preparo e Pré-processamento dos Dados, e com o não-aproveitamento de todas variáveis utilizadas no modelo após a redução do espaço de input.

Embora não tenha sido o escopo deste trabalho, ficou claro que a extração dos dados de output deste modelo, de forma automática ou manual, permitirá a otimização das Campanhas de Email Marketing, e até de outras plataformas, caso o comportamento Omnicanal “Email Marketing – loja física” se estenda aos outros canais digitais. Levará também à melhora contínua do próprio modelo, a partir das análises de performance das Campanhas.

8. LIMITAÇÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Algumas limitações a serem destacadas durante este estudo foram a obtenção de dados exclusivamente pela empresa Fnac Portugal – a recolha de dados da Fnac em outros países Europeus, ou mesmo de outras empresas do setor do Retalho, permitiria a criação de um modelo mais sofisticado; o fato de o estudo ter sido realizado apenas com dados de uma Campanha, o que impossibilitou a realização de análises comparativas e a confirmação das tendências observadas.

Faz-se necessário que este estudo, e o modelo dele decorrente, sejam revisados por pares académicos e aperfeiçoado quanto às técnicas de Data Mining utilizadas. Outros algoritmos devem ser testados dentro da mesma perspetiva. A automatização das etapas deste modelo, bem como a de recolha de dados de outros canais digitais, principalmente dados não-estruturados, dentro da estratégia Omnicanal, compõem uma linha promissora de pesquisa.

9. BIBLIOGRAFIA

- Beaujean, M., Davidson, J., & Madge, S. (2006). The “moment of truth” in customer service. *Mckinsey Quarterly*, 1, 62-73.
- Becker, D. O. (2003). Gambling on customers. *McKinsey Quarterly*, (2), 46-153.
- Berry, M.J.A. & Linoff, G. (2004). *Data mining techniques: for marketing, sales, and customer support*. (2nd ed.) Indianapolis: Wiley Publishing, Inc.
- Berry, M. J., & Linoff, G. S. (2006). *Mastering data mining: the art and science of customer relationship management*. Amazon.
- Blattberg, R. C., Kim, B. D., & Neslin, S. A. (2008). *Why Database Marketing?* (pp. 13-46). Springer New York.
- Blattberg, R. C., Kim, B. D., & Neslin, S. A. (2010). *Database Marketing: Analyzing and Managing Customers* (Vol. 18). Springer Science & Business Media.
- Davenport, T. H., & Dyché, J. (2013). *Big data in big companies*. International Institute for Analytics.
- Dyché, J. (2002). *The CRM handbook: a business guide to customer relationship management*. Addison-Wesley Professional.
- Dyché, J., & Levy, E. (2011). *Customer data integration: Reaching a single version of the truth* (Vol. 7). John Wiley & Sons.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis. *Psychological assessment*, 12(1), 19.
- Han, Jiawei, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.
- Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. MIT press.
- Imhoff, C., Loftis, L., & Geiger, J. G. (2001). *Building the customer-centric enterprise: Data warehousing techniques for supporting customer relationship management*. Wiley.
- Kamakura, W., Mela, C. F., Ansari, A., Bodapati, A., Fader, P., Iyengar, R., ... & Wedel, M. (2006). Choice models and customer relationship management. *Marketing Letters*, 16(3-4), 279-291.
- Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (2013). *Machine learning: An artificial intelligence approach*. Springer Science & Business Media.
- Pyle, D. (1999). *Data preparation for data mining* (Vol. 1). Morgan Kaufmann.
- Richards, K. A., & Jones, E. (2008). Customer relationship management: Finding value drivers. *Industrial marketing management*, 37(2), 120-130.

Rigby, D. 2011. The future of shopping. *Harvard Business Review*, December 2011.

Shankar, V., Inman, J. J., Mantrala, M., Kelley, E., & Rizley, R. (2011). Innovations in shopper marketing: current insights and future research issues. *Journal of Retailing*, 87, S29-S42.

Siddiqi, J. I., Akhgar, B., Wise, T., & Hallam, S. (2006). A Framework for the Implementation of a Customer Relationship Management Strategy in Retail Sector. In *IKE* (pp. 244-249).

Sofia, H. & Henriques, R. (2013). Building clusters for CRM strategies by mining airlines customer data". *Information Systems and Technologies (CISTI)*, 2013 8th Iberian Conference on, Lisbon.

Tseng, M. M., & Hu, S. J. (2014). Mass customization. In *CIRP Encyclopedia of Production Engineering* (pp. 836-843). Springer Berlin Heidelberg.

Weiss, S. M., & Indurkha, N. (1998). *Predictive data mining: a practical guide*. Morgan Kaufmann.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Wojciech Piotrowicz and Richard Cuthbertson, Guest Editors *International Journal of Electronic Commerce*, Volume 18, Number 4, Summer 2014, pp. 5-16.

10.ANEXOS

10.1. ANEXO A: EXEMPLO DE EMAIL MARKETING DNA FNAC

CONVITE ESPECIAL PARA CLIENTES ESPECIAIS

3 e 4 março

DIAS ADERENTE FNAC

Faz a festa

ATÉ **40%**

DESCONTO IMEDIATO

TECNOLOGIA | LIVROS | ENTRETENIMENTO

EVENTOS AO VIVO

Ver todos os eventos dos Dias Aderente FNAC

PARA TER O CARTÃO FNAC SEMPRE À MÃO....

APP CARTÃO FNAC

SÓ TENS DE MEXER UM DEDO

VALES ADERENTE | CAMPANHAS FNAC | O MEU SALDO

The banner features a bright yellow background with a white checkered pattern in the top right corner. It includes images of a laptop, a blue bag, and a game controller. The bottom section has a black background with a microphone icon and a grid of six small images showing various people and a cartoon character. The bottom right of the banner shows a woman smiling while using a smartphone.

Figura 7 – Peça de Email Marketing DNA Fnac de março de 2017. <http://www.fnac.pt>

10.2. ANEXO B: DIAGRAMA ANALÍTICO DO PROJETO EM SAS ENTERPRISE MINER

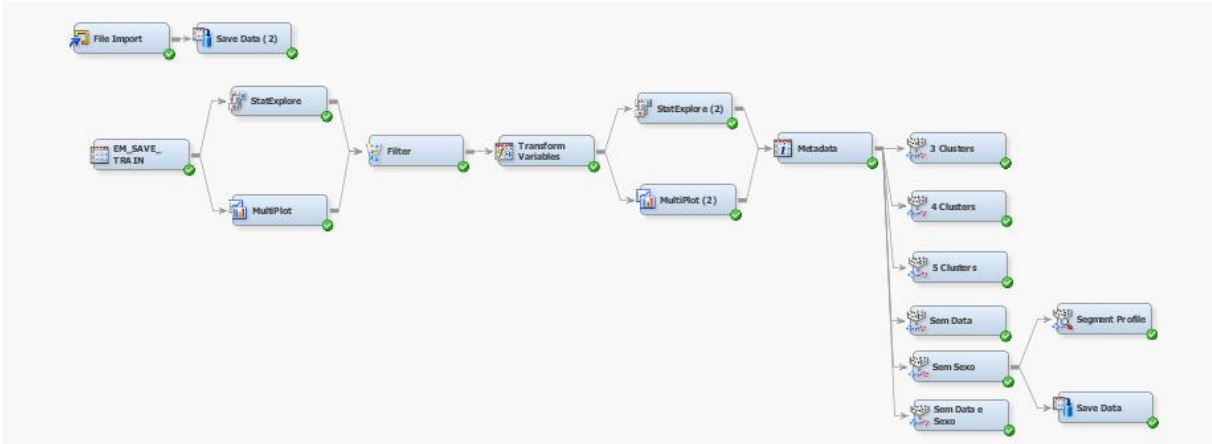


Figura 8 – Diagrama Analítico do modelo de Segmentação. SAS Enterprise Miner 9.4