

Escola de Comunicação, Artes e Tecnologias da Informação



# **ANÁLISE DO PERFIL DE CLIENTES COM POSSE DE PRODUTOS DE INVESTIMENTO NA BANCA**

---

Marcelo Filipe Santos de Pinho

Aluno nº 20064871

Trabalho Final de Curso  
Licenciatura em Engenharia Informática

Orientadora:

Prof. Mestre Ana Margarida Gomes Alexandre, Data Mining, Informática de  
Gestão

Julho 2011

# Índice

<b>Índice de Figuras .....</b>	<b>4</b>
<b>Índice de Tabelas .....</b>	<b>4</b>
<b>Resumo.....</b>	<b>5</b>
<b>Abstract .....</b>	<b>6</b>
<b>1. Introdução .....</b>	<b>7</b>
1.1 Relevância do Modelo de Segmentação de Clientes com posse de produtos de investimento na Banca .....	7
1.2 Objectivos e Descrição do Trabalho Final de Curso .....	8
<b>2. Enquadramento Teórico .....</b>	<b>9</b>
2.1 Gestão da Relação com o cliente .....	10
2.2 Modelos de Data Mining como processos relevantes no conhecimento dos Clientes ..	11
2.3 O processo de Segmentação na era passada.....	12
<b>3. Método .....</b>	<b>13</b>
3.1 Introdução .....	13
3.2 Metodologia .....	13
3.3 Dados Utilizados .....	15
3.4 Produtos utilizados no desenvolvimento do modelo .....	16
3.5 Processo de desenvolvimento do Modelo .....	16
<b>4. Resultados.....</b>	<b>20</b>
4.1 Análise dos Segmentos Criados .....	20
4.2 Oportunidades/Campanhas sobre cada um dos Segmentos .....	22
<b>5. Conclusões e Trabalho Futuro .....</b>	<b>23</b>
5.1 Conclusões .....	23
5.2 Trabalho Futuro .....	24
<b>Bibliografia .....</b>	<b>25</b>
<b>Anexos.....</b>	<b>27</b>
Anexo 1. Variáveis que constam na ABT de desenvolvimento.....	27
Anexo 2. Código SAS- Pré-tratamento de Missings .....	28
Anexo 3. Nó Impute e as respectivas propriedades associadas .....	28
Anexo 4. Nó Filter e as respectivas propriedades associadas .....	28

Anexo 5. Nó Transform Variables com aplicação do método Bucket às variáveis intervalares	28
Anexo 6. Nó Cluster com o método Ward e o respectivo perfil de segmentos .....	29
Anexo 7. Nó Cluster com o método K-Means com a definição do número de clusters a ser construído e o respectivo perfil de segmentos .....	29
Anexo 8. Variáveis mais relevantes para a solução dos Segmentos obtidos .....	29
Anexo 9. Caracterização dos Segmentos.....	30
Anexo 10. Dimensão dos Segmentos .....	31
Anexo 11. Variáveis mais relevantes para cada um dos Segmentos obtidos.....	32
Anexo 12. Comparação das características dos Segmento face às características da População	32
Anexo 13. Source Code do Projecto .....	34

## Índice de Figuras

Figura 1. Metodologia de Plataformas Inteligentes (modelo adaptado à realidade deste projecto).....	14
Figura 2. Diagrama que reflecte a criação do Modelo de Segmentação.....	17
Figura 3. Distância entre Clusters .....	20

## Índice de Tabelas

Tabela 1. Resumo dos Perfis de cada Segmento .....	21
Tabela 2. Oportunidades para cada um dos Segmentos .....	22

## Resumo

Hoje em dia as entidades financeiras vivem uma situação económica crítica a nível mundial e com uma forte concorrência. As entidades financeiras tiveram que se reajustar ao panorama económico actual, de forma que conseguissem sobreviver no mercado. A gestão da relação com o cliente passou a ser fundamental para o sucesso da empresa. Foi com o propósito de conhecer melhor os clientes (características e comportamentos) que se desenvolveu e implementou o modelo de segmentação de clientes com posse de produtos de investimento na banca. Este modelo foi desenvolvido e implementado de acordo com a metodologia (IPI), “Metodologia de Plataformas Inteligentes” que é utilizada pelo SAS para o desenvolvimento de projectos Data Mining. O modelo desenvolvido permitiu agrupar clientes de acordo com as características e padrões em segmentos, bem como efectuar uma análise de perfil. Para cada segmento foram desenvolvidas um conjunto de estratégias de marketing adequadas a cada perfil.

**Palavras – Chave:** Segmentação, Mineralização de Dados, Gestão de Relacionamento com o Cliente

## **Abstract**

Nowadays financial institutions live a critical economic situation worldwide and the strong competition. Financial institutions had to adjust to the current economic outlook so they could survive. The customer's relationships management has become critical to business success. It was in order to better know the client (characteristics and behaviors) that developed and implemented a customer segmentation model with ownership of investment products in banking. This model was developed and implemented in accordance with the methodology (IPI), "Methodology for Intelligent Platforms" that is utilized by SAS for the development of data mining projects. With the develop model was possible join customers with common characteristics and patterns. For each cluster were established appropriate marketing strategies to each segment profile.

**Keywords:** Cluster, Data Mining, Customer Relationship Management

# 1. Introdução

Segundo o estudo e teorias de (Woodcock, 2002) a extracção e posterior análise das características inerentes, desejos e necessidades dos clientes no sector bancário, cria um relacionamento, de ambas as partes, positivo na medida em que se conquista a confiança do cliente. Por outro lado, as organizações financeiras necessitam de ter um conhecimento mais profundo e eficiente sobre cada um dos clientes, de modo a desenvolver interacções positivas e consistentes que visam a oferta adequada de produtos e soluções. Os processos desenvolvidos e que visem intensificar e melhorar o relacionamento com os clientes no curto, médio e longo prazo utilizado pelos bancos têm contribuído para a satisfação, retenção e consequente fidelização dos clientes, de uma forma mais eficaz e fulcral para o sucesso da instituição bancária. Estes processos têm vindo a ganhar cada vez maior importância e relevância.

Dado à conjectura económica actual sentida no país e à concorrência existente entre as entidades bancárias, estas viram-se obrigadas a redesenhar as suas estratégias de negócio para atrair novos clientes<sup>1</sup> e para reter os clientes já existentes<sup>2</sup>. Isto porque, custa mais captar um novo cliente do que reter clientes já existentes (Sheth, 2001). Contudo, todos estes aspectos são cada vez mais complexos e difíceis de se obter, já que os clientes são cada vez mais aliciados, agressivamente, pela concorrência e estão cada vez mais exigentes.

## 1.1 Relevância do Modelo de Segmentação de Clientes com posse de produtos de investimento na Banca

A dificuldade (ou mesmo ausência) de comunicações direccionadas, eficazes e adequados ao perfil de cada cliente investidor e a facilidade com que cliente poderá deixar de possuir produtos de investimento, são factores de alerta que levaram as entidades bancárias a pretenderem conhecer as características e necessidades dos seus clientes. Surge então o Modelo de Segmentação de Clientes com posse de Produtos de Investimento, algo que tem vindo a ganhar cada vez mais relevância.

Dado à natureza e características de diferentes produtos de investimento existentes no mercado é fulcral medir o retorno esperado perante a oferta variada de produtos de investimento. O investimento é classificado por três características distintas: risco/volatilidade, retorno e liquidez (Macedo, 2010).

---

<sup>1</sup> Especialmente clientes com um perfil cada vez mais investidor e de maior retorno financeiro.

<sup>2</sup> Especialmente clientes de maior valor e cujo perfil é investidor.

Mediante estas características e condicionantes nos produtos de investimento as organizações tiveram de se adaptar a uma realidade estratégica de negócio distinta. A existência de um grande conhecimento dos clientes, consoante as suas características comportamentais e sociodemográficas, é o melhor ponto de arranque para segmentar o mercado (Kotler, 1998).

## **1.2 Objectivos e Descrição do Trabalho Final de Curso**

Os Modelos de Data Mining<sup>3</sup> assumem um papel relevante no processo de gerir a relação com os clientes. Este trabalho final de curso teve como principal objectivo o desenvolvimento de um Modelo de Data Mining Comportamental, mais concretamente de um Modelo de Segmentação de Clientes com posse de produtos de investimento na Banca. O objectivo é criar, descrever e analisar o perfil de cada grupo de clientes<sup>4</sup>. Para tal, iremos utilizar como input características sociodemográficas do próprio cliente e características inerentes à posse de produtos de investimento. Todas estas características<sup>5</sup> permitem avaliar o cliente no seu todo e desenvolver um conjunto de acções/campanhas de marketing direccionadas segundo cada um dos perfis existentes.

---

<sup>3</sup> Os Modelos de Data Mining dividem-se em dois grandes grupos: (1) Modelos Preditivos e (2) Modelos Comportamentais. Os Modelos Preditivos traduzem um conjunto de modelos cujo objectivo é prever e explicar um determinado evento. Já os Modelos Comportamentais dizem respeito a modelos/técnicas cujo objectivo é descrever e analisar o comportamento dos clientes com base no seu envolvimento passado e actual.

<sup>4</sup> O Modelo de Segmentação é um método de Aprendizagem Não Supervisionado e traduz o processo de agrupar observações (clientes) em grupos o mais heterogéneos entre si e o mais homogéneos dentro do mesmo grupo.

<sup>5</sup> Características intrínsecas e características que traduzem o perfil de envolvimento com a instituição.



## 2. Enquadramento Teórico

O crescimento da informação nos últimos vinte anos foi de tal forma grande que originou o desenvolvimento de sistemas, com base nas tecnologias da informação que permitissem a recolha, processamento, transmissão e disseminação de dados com valor para os utilizadores. Estes sistemas são conhecidos como Sistemas de Informação (SI). A razão mais forte do desenvolvimento e implementação destes Sistemas pelas organizações é essencialmente para resolver problemas organizacionais e para reagir à mudança no meio envolvente (C. Laudon, 2000).

Dado à globalização, os sistemas de informação passaram a assumir um papel de extrema importância nas organizações. Quando estes sistemas passam a concentrar toda a informação e dar um grande apoio na gestão das empresas passaram a ser denominados de Sistemas de Informação de Gestão (O'Brien, 2006). Um desenvolvimento de gestão eficaz e eficiente pressupõe em qualquer organização a existência de um bom sistema de informação para a tomada de decisão de forma ágil e segura. Na actualidade e na sociedade da informação em que se vive, é exigido às organizações planos de gestão estratégica eficientes que são facilmente conseguidos com a facilidade do uso das tecnologias da informação. Nenhuma organização consegue ter vantagem competitiva sem uso destas ferramentas (Porter, 1985).

Os gestores de topo têm a necessidade de tomar decisões de uma forma rápida, flexível, dinâmica e eficaz, razão pela qual surgiram os Sistemas de Suporte à Decisão (SSD). Assim, os SSD desenham e assistem a decisão táctica e estratégica orientada ao assunto, recorrendo à análise de um conjunto de dados históricos de múltiplas fontes (Turban, 1998). Os SSD são sistemas de informação desenvolvidos para níveis de gestão das empresas<sup>6</sup>, que adaptam dados e modelos analíticos sofisticados e técnicas de Data Mining (DM) para o suporte da tomada de decisão. Estes Sistemas devolvem resultados e conclusões sobre a forma de Relatórios com uma representação gráfica bastante intuitiva capazes de ajudar o processo de tomada decisão.

Os SSD são classificados em SSD Formal e SSD Ad-Hoc, com base nos atributos do contexto de resolução de problemas (Madnick, 1977).

O SSD Formal foca-se no tipo de decisões periódicas ou recorrentes dentro da organização.

---

<sup>6</sup> Responsáveis pelo planeamento táctico estratégico na organização.

Os SSD ad-hoc são construídos para responder a contextos de problemas muito restritos, a conjuntos de decisões não recorrentes ou de fácil antecipação. Um cenário típico de utilização deste tipo de SDD seria uma fusão agressiva entre empresas.

Com o desenvolvimento das novas tecnologias, estes sistemas estão a tornar-se menos dispendiosos, com a vantagem de terem elevada eficácia. Com a evolução dos SDD, as organizações passaram a ter uma actuação on-line Business to Consumer (B2C) e com esta medida surgiram meios de interacção/canal de comunicação com o cliente/consumidor que outrora não existia antes do aparecimento dos SI. Este SSD, com suporte de gestão de relacionamento com os clientes, passou a ser referido de Gestão da Relação com o Cliente (Customer Relationship Management).

## 2.1 Gestão da Relação com o cliente

As organizações de hoje em dia precisam de gerir os seus recursos de forma mais eficiente e mais orientada às necessidades do cliente, ou seja, quanto mais perto as organizações estiverem do cliente melhor a gestão dessa relação (Berry, 1983). Esta nova forma de interacção com o cliente permite reter informações úteis sobre os mesmos, com o intuito de compreender as suas características, necessidades e satisfazê-los em todo o ciclo de vida dos mesmos na organização. Com as tecnologias da informação o *Customer Relationship Management* (CRM) é descrito como sendo uma forma de relação focada individual, cliente a cliente, que integram a base de dados de conhecimento com uma estratégia de crescimento do negócio na retenção e fidelização de clientes a longo prazo (Peppers D. e Rogers, 1993).

Existem três tipos de CRM: CRM Operacional, CRM Analítico e CRM Colaborativo (METAGroup, 2000). Os conceitos de CRM Operacional e CRM Analítico são dois conceitos muito próximos mas que definem diferentes aspectos do CRM (Dyche, 2003). O CRM Operacional relaciona factores operacionais na implementação do sistema de CRM e permite integrar funções relacionadas com o serviço do cliente<sup>7</sup>.

O CRM Analítico está relacionado com a análise e interpretação de dados (estatística e de data mining) armazenados dos clientes. Esta análise e exploração dos dados armazenados pela área operacional é o processo de CRM Analítico.

---

<sup>7</sup> Tais como: gestão de pedidos, sistemas de facturação, gestão e automatização do mercado e força de vendas

Por fim e não menos importante o CRM Colaborativo. Este tipo de CRM reúne todos os canais de contacto com o cliente (email, fax, cartas, contacto telefónico). Estes canais de contacto devem garantir a interacção dos dados com toda a organização.

Este trabalho baseia-se no desenvolvimento de um processo de CRM Analítico, onde foram visualizados, manipulados e desenvolvidas um conjunto de análises aos dados dos clientes através de Modelos de Data Mining.

## **2.2 Modelos de Data Mining como processos relevantes no conhecimento dos Clientes**

Desde os anos noventa que a mineralização de dados é visto como o processamento de dados proveniente de diversos domínios da indústria desde a área académica, empresarial até às ciências da saúde (Gorunescu, 2011). Do ponto de vista da investigação, o Data Mining (DM) ainda é visto como um tema pouco desenvolvido pois é muito recente historicamente. O DM traduz o processamento de selecção, exploração, análise e modelação de grandes bases de dados, recorrendo a diversas técnicas comportamentais e preditivas. O principal objectivo é analisar tendências, padrões, relações, associações entre os dados, ou seja, transformar dados em informação e a informação em conhecimento. Tudo isto permite uma extracção não trivial de conhecimento e potencialmente útil para as empresas (Adriaans, 1996).

Este conhecimento é uma mais-valia para o planeamento, gestão e tomada de decisão, mas, no entanto, por vezes difícil de analisar. Isto porque, muitas das vezes, esta informação está implícita, escondida ou mesmo dispersa num grande volume de dados.

O processo de Data Mining envolve essencialmente três passos muito importantes (Gorunescu, 2011):

- Exploração de dados<sup>8</sup> - Esta fase envolve limpeza<sup>9</sup>, modificação, transformação, dimensionamento e selecção de uma parte dos dados;
- Criação do modelo - são desenvolvidos vários modelos, segundo diversas técnicas e parametrizações, e procede-se à comparação dos mesmos, de acordo com um conjunto de critérios. O objectivo é identificar e avaliar/ajustar o melhor modelo;
- Pós-processamento e Capacidade de Generalização do Modelo - traduz a fase final do processo e que diz respeito à aplicação do modelo a novos dados para garantir que o modelo desenvolvido se aplica a novos dados.

---

<sup>8</sup> Exploração de dados em bruto e de grande volume.

<sup>9</sup> Data Cleaning é efectuada no pré-processamento de dados que visa assegurar a qualidade dos dados.

Tipicamente, os modelos de segmentação são modelos considerados chave no processo de Data Mining, uma vez que permitem identificar e conhecer as características e comportamentos dos clientes (algo que constitui a base de qualquer tipo de interacção). Assim sendo, a Segmentação é um princípio fundamental no marketing de relacionamento (Kotler, 1998). Este modelo traduz o processo de classificar clientes em grupos heterogéneos entre si e em grupos homogéneos (sendo que os clientes associados a um determinado segmento partilham um perfil similar).

Do ponto de vista da gestão da relação com o cliente, clientes segmentados permitem às empresas criar e diferenciar produtos que visem a satisfação total do cliente de cada perfil tendo como retorno a sua lealdade (Kotler, 1998). Para que exista esta lealdade as empresas têm que ter sempre o cuidado em gerar valor acrescentado ao produto/serviço prestado.

### **2.3 O processo de Segmentação na era passada.**

Numa abordagem tradicional, o processo de segmentação era baseado em campanhas de Marketing Massivas e não direccionadas. Assim, as empresas ignoravam a possibilidade de desenvolverem acções de marketing personalizadas a cada cliente, principalmente devido à focagem das mesmas em produzir, distribuir e promover um só produto para chegar a todo o mercado. O contacto com o cliente era somente baseado no produto. No passado, este tipo de marketing massivo assumia um grande potencial de mercado, uma vez que não existia muita concorrência, os consumidores não eram muito exigentes e viam os bancos como entidades superiores. Estes motivos levavam a que os custos fossem menores, as margens de lucro maiores e houvesse um potencial de mercado (Armstrong/Kotler, 2003).

Actualmente as acções/campanhas de marketing efectuadas sobre os clientes deverão ser, tipicamente, uma abordagem One-to-One por segmento, uma vez que maior é a concorrência e exigência dos consumidores. Desta forma, consegue-se desenvolver acções de comunicação e marketing personalizadas, gerando um melhor retorno às empresas.

Futuramente, é expectável, e tendo em conta a evolução dos SI's, que as campanhas de marketing que sejam exclusivas a cada cliente (e não analisando o perfil do segmento em que o mesmo se insere). Assim, o intuito será analisar as características cada cliente de uma forma individual e oferecer um determinado produto/serviço que se adapta na perfeição ao cliente.

## 3. Método

### 3.1 Introdução

Este projecto teve como principal objectivo segmentar clientes bancários com posse de produtos de investimento, com o intuito de identificar e analisar o perfil de risco. Assim, este estudo permitiu identificar e compreender se um determinado cliente tem um comportamento conservador, arriscado ou uma carteira de produtos variada.

Os dados utilizados neste projecto são dados reais provenientes de uma instituição bancária, que por motivos confidenciais não será divulgada.

O modelo desenvolvido e implementado é o modelo de desenvolvimento de Clusters<sup>10</sup> que se baseia em conjugar observações em grupos o mais distintos entre si e o mais similares dentro do mesmo grupo. O objecto de estudo resumiu-se num conjunto de clientes do retalho bancário com as seguintes características: (1) clientes particulares; (2) clientes não colaboradores; (3) clientes no banco há mais de seis meses; (4) clientes activos no mês actual.

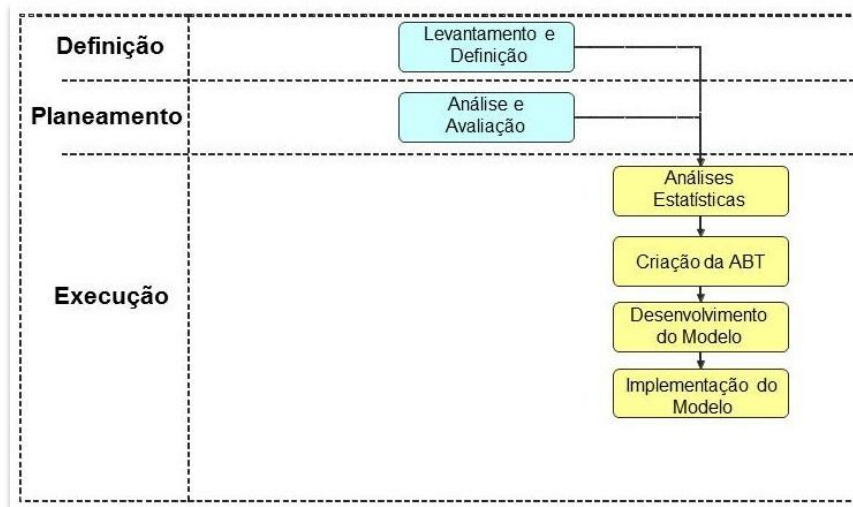
### 3.2 Metodologia

A metodologia utilizada na criação deste projecto é a metodologia da SAS que está vocacionada para o desenvolvimento de projectos – “Metodologia de Plataformas Inteligentes” (Intelligence Platform Methodology) (IPI). Esta metodologia é composta por seis etapas distintas em funções e características (Figura 1).

---

<sup>10</sup> A Construção de um Cluster envolve o processo estatístico de partição do universo de dados por grupos. A construção de um Cluster desenvolve-se por dois métodos: O Método Hierárquico ou o Método Não-hierárquico. (Kumar, 2005). Neste projecto em particular a construção de clusters será pelo Método Não-Hierárquico que requer como input o número de clusters a serem formados.

Figura 1. Metodologia de Plataformas Inteligentes (modelo adaptado à realidade deste projecto)



Fonte: SAS® Intelligence Platform Implementation Methodology by SAS Institute Inc., Cary, NC, USA, 2006

- Primeira Etapa: Levantamento e definição

Foi feita uma avaliação da situação actual do cliente no banco e uma definição clara dos objectivos do modelo a desenvolver de modo a evitar um planeamento desadequado.

- Segunda Etapa: Análise e Avaliação

Esta etapa consistiu numa análise à disponibilidade dos dados existentes e adequações. Foram analisados dados históricos de clientes referentes a 12 meses (entre Março de 2010 e Março de 2011), dando enfoque aos últimos três, seis, doze meses e mês actual<sup>11</sup>. Para cada momento temporal foram extraídos os dados pelo banco para a posterior exploração dos mesmos.

Efectuou-se, ainda, uma avaliação do software a utilizar para o desenvolvimento do modelo de segmentação. Quanto à escolha do Software optou-se pelo:

- SAS Enterprise Guide: manipulação e transformação de dados;
- SAS Enterprise Miner: desenvolvimento e implementação do modelo comportamental de segmentação.

<sup>11</sup> Mês actual refere-se a Março de 2011

- Terceira Etapa: Análises Estatísticas

Após a análise e à disponibilidade dos dados foi efectuada uma análise estatística preliminar aos dados provenientes de diferentes Sistemas de Bases de Dados Relacionais (SQL Server, Oracle) para aferir a qualidade e possíveis adequações dos mesmos. Foi ainda realizada toda uma preparação dos dados para depois serem englobados na *Analytical Base Table*.

- Quarta Etapa: Criação da *Analytical Base Table* (ABT)

*Analytical Base Table* é uma tabela que é criada e usada para o desenvolvimento e aferição de modelos analíticos e de previsão de comportamentos dos clientes. Cada registo existente na ABT representa um cliente e contém os dados que descrevem o seu comportamento. A ABT utilizada no desenvolvimento do Modelo tem 63 variáveis e 64193 Clientes.

- Quinta e Sexta Etapa: Desenvolvimento e Implementação do Modelo

Nesta etapa foi desenvolvido o modelo comportamental da Segmentação. Este modelo foi desenvolvido através de um processo denominado SEMMA<sup>12</sup>, onde foi possível extrair, explorar, modelar, modificar e avaliar os dados de forma a criar segmentos de clientes com posse de produtos de investimento.

### 3.3 Dados Utilizados

O Modelo Comportamental da Segmentação foi desenvolvido para 64193 clientes activos, particulares, não falecidos, não colaboradores, com escalão de antiguidade superior a 6 meses e clientes cujo somatório dos saldos em Depósitos a Prazo, Fundos, Poupanças, Produtos Fiscais, Seguros no mês actual foi superior a 250 €.

É importante salientar que as variáveis descritivas são todas aquelas que, embora não sejam componentes da construção do cluster, irão servir para analisar o perfil dos segmentos. Antes de iniciar o desenvolvimento do modelo em causa procedeu-se à análise preliminar de algumas variáveis sociodemográficas, tais como: sexo, faixa etária, imigrante e nacionalidade. Verificou-se que estes clientes são maioritariamente do sexo masculino (60,33%), entre os 40 e os 55 anos (35,62%), maioritariamente de

---

<sup>12</sup> SEMMA é acrónimo de Sample, Explore, Model, Modify, Asset pela SAS Institute Inc.

Nacionalidade Portuguesa (95,76%) e minoritariamente imigrantes (5,60%). Estes clientes apresentavam em produtos de Alto Risco nos últimos 12 meses um montante médio de 3384668 euros e com um montante máximo de 2502282 euros. Em relação as mesmas estatísticas mas referentes a Produtos de Baixo Risco nos últimos 12 meses apresentavam um montante médio de 6561443 euros tendo no máximo 1832997 euros em produtos desta natureza.

### **3.4 Produtos utilizados no desenvolvimento do modelo**

Os produtos utilizados para o desenvolvimento do Modelo da Segmentação foram os seguintes:

- Depósitos a Prazo (D\_PRAZO);
- Poupanças (POUP);
- Produtos Fiscais (P\_FISCAIS);
- Seguros de Capitalização (S\_CAP);
- Fundos (FU);
- Produtos Estruturados (PE).

Tendo em consideração a avaliação de risco pelo Banco, este identifica como produtos de baixo risco (PBR): D\_PRAZO, POUP, P\_FISCAIS e S\_CAP e produtos de alto risco (PAR): FU e PE. Esta divisão clara entre PBR e PAR é um processo inicial para compreender em que forma é que os clientes são amantes do risco ou avessos ao risco. Dentro de uma diversidade de produtos, foram estes os produtos escolhidos pelo banco uma vez que originam maior retorno.

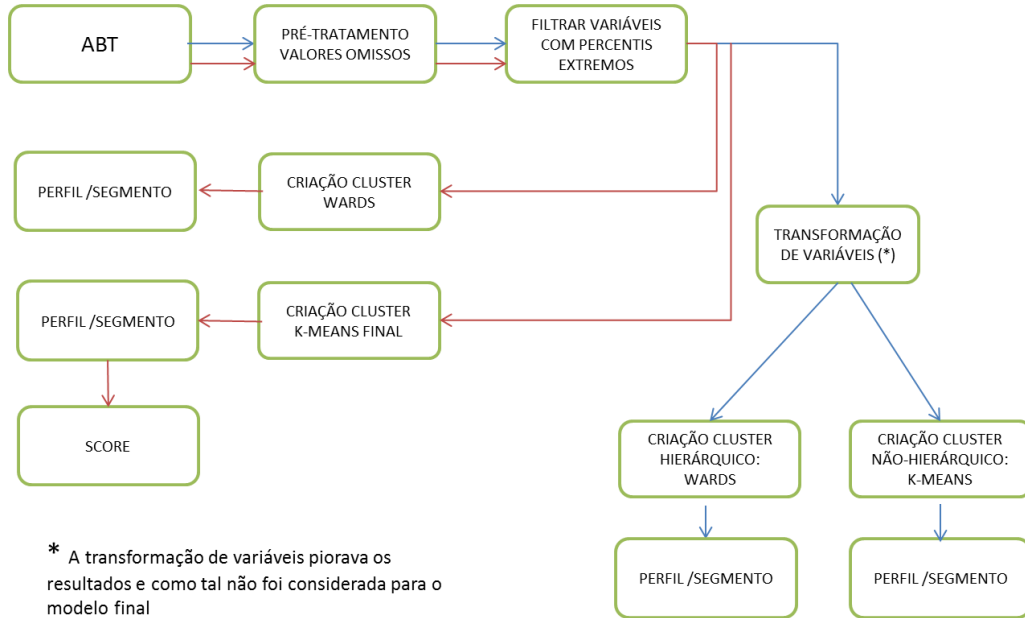
### **3.5 Processo de desenvolvimento do Modelo**

O processo de desenvolvimento do modelo envolveu um conjunto de etapas. A figura 2 ilustra o processo criado no SAS Enterprise Miner. Cada uma destas etapas teve objectivos distintos e foram desenvolvidas com o intuito de explorar os dados, proceder ao tratamento de problemas dos dados, modificar dados, modelar os dados e implementar o modelo desenvolvido periodicamente. De realçar que foram criados



dois cenários distintos, um que engloba transformações de variáveis (etapas representadas a azul) e outro cenário sem estas (etapas representadas a vermelho).

Figura 2. Diagrama que reflecte a criação do Modelo de Segmentação



Analisemos então cada uma das etapas, sequencialmente, realizadas no desenvolvimento deste modelo:

- (1) **ABT:** Numa primeira fase procedeu-se à importação da ABT a ser utilizada no desenvolvimento do modelo e a toda uma adequação da metadata das variáveis a serem utilizadas. Foi então definido qual o tipo e papel que cada uma das variáveis assumiria no modelo a desenvolver. (Anexos – Anexo 1)
- (2) **Pré-Tratamento dos Valores Omissos:** procedeu-se ao pré-tratamento dos valores omissos nas variáveis, uma vez que existia um conjunto de variáveis que apresentavam valores omissos. Este pré-tratamento das variáveis foi efectuado através de código SAS<sup>13</sup> (onde se atribuiu um determinado valor conhecido aos valores omissos de algumas variáveis) e através do nó de Impute<sup>14</sup>. Através deste nó foi possível atribuir a cada missing um valor constante, tanto para as variáveis categóricas (onde foi criada uma nova classe nestas variáveis, classe esta que reflectia tudo o que fosse desconhecido (“D”)) como para as variáveis

<sup>13</sup> Anexo 2 – Código SAS Pré-tratamento de Valores Omissos.

<sup>14</sup> Funcionalidade existente no SAS Enterprise Miner e cujo objectivo é realizar o tratamento dos valores omissos através de métodos estatísticos ou constantes definidas (Anexo 3).

intervalares (onde substituímos cada valor omissso pelo valor “0”). De realçar que este tipo de procedimento é, tipicamente, aconselhado para a análise de clusters pois o resultado poderá traduzir-se na simplificação da interpretação dos segmentos obtidos.

**(3) Filtro de Variáveis:** Tendo em conta que existiam variáveis intervalares com valores muito dispersos e acentuados aplicou-se o nó de filtro<sup>15</sup>. Este nó do SAS Enterprise Miner permite filtrar variáveis, utilizando para tal diversos métodos estatísticos (com objectivos distintos). No trabalho em causa aplicou-se como filtro o método dos Percentis Extremos, o que significa que foram filtrados valores acima dos percentis extremos.

**(4) Transformação de Variáveis:** Nesta etapa verificou-se a necessidade de transformar as variáveis intervalares em variáveis de classe, a fim de agrupar os dados em intervalos. Contudo, e após alguns testes de transformações distintas (entre elas o método de Bucket<sup>16</sup>), os resultados finais ficavam muito aquém do esperado e por esse motivo resolveu-se não considerar as transformações de variáveis para o modelo em causa.

**(5) Criação dos Segmentos:** Foi então aplicado após o nó de filtros um nó Clusters. Através deste nó foi possível desenvolver dois tipos de Segmentos: (1) Clusters Hierárquicos e (2) Clusters Não Hierárquicos.

- Clusters Hierárquicos: Segundo (Sokal, 1973), a construção de Clusters Hierárquicos implica agrupar observações com elevada similaridade entre elas e segundo vários critérios de aglomeração<sup>17</sup> que são os seguintes: (1) Single Linkage<sup>18</sup>, (2) Complete Linkage<sup>19</sup>, (3) Average Linkage<sup>20</sup>, (4) Centroide<sup>21</sup> e (5) Ward<sup>22</sup>. Neste modelo em causa foi inicialmente aplicado o nó de Cluster com o método Ward<sup>23</sup>.

---

<sup>15</sup> Anexo 4 - Nó Filter.

<sup>16</sup> Anexo 5 – Nó Transform Variables com o método Bucket (agrupamento de dados em intervalos).

<sup>17</sup> Aglomeração de observações com algoritmos de agrupamento

<sup>18</sup> Single Linkage é a distância entre dois clusters mínima ou seja, é a distância entre dois pares de objectos mais próximos.

<sup>19</sup> Complete Linkage é a distância entre dois clusters máxima entre eles, isto é a distância entre dois pares de objectos mais distantes.

<sup>20</sup> Average Linkage é a distância média entre dois clusters, como sendo a distância média entre dois pares de objectos

<sup>21</sup> Centroid é a distância entre dois clusters que é definida como a distância entre centroides(ponto médio num espaço multidimensional definido pelas variáveis) dos dois clusters.

<sup>22</sup> Ward Linkage Utiliza a análise da variância para avaliar a distância entre clusters. Esta regra tenta minimizar a soma dos quadrados de quaisquer dois clusters que poderão dar a origem a um único cluster. Eficaz mas tende a criar grupos pequenos.

<sup>23</sup> Anexo 6 – Nó Cluster com o método Ward e o respectivo perfil de segmentos

- **Clusters Não Hierárquicos:** são clusters que requerem, à partida, a definição do número de clusters a serem formados pelo próprio utilizador. Assim, este método é assim mais rápido e exige um menor esforço de processamento. Neste trabalho a segmentação não hierárquica foi desenvolvida de acordo com o método **K-Means**<sup>24</sup>.

Para a criação dos segmentos utilizou-se, numa primeira fase, uma Segmentação Hierárquica (utilizando o método de Ward). Como resultado obteve-se 10 segmentos (k=10). Contudo, a após uma análise à matriz de distâncias, verificou-se que alguns dos segmentos apresentavam distâncias (e como tal semelhanças) muito próximas. Assim, e numa segunda fase, realizou-se uma Segmentação Não Hierárquica, através do método de K-Means, onde o objectivo foi a redução, de forma iterativa (e descendentemente), do número de clusters obtido inicialmente (testou-se k=10,9,8, etc). A solução óptica obtida foi de 4 segmentos, uma vez que estes se encontravam bem distanciados uns dos outros e como tal o grau de dissemelhança era máximo entre os objectos dos clusters.

**(6) Análise de Perfil dos Segmentos:** De seguida foi adicionado o nó de “Segment Profile”. Este nó permitiu: (1) analisar o perfil dos segmentos de acordo com outras variáveis não utilizados no desenvolvimento dos segmentos<sup>25</sup>; (2) identificar as variáveis mais relevantes para cada segmento e (3) comparar as características de cada segmento com as características da população.

**(7) Implementação do Modelo:** Por último, e com o intuito de se obter todo o código subjacente à criação do modelo para a scorização automática e periódica do modelo, foi aplicado nó de Score. Este nó, existente no SAS Enterprise Miner, permite aplicar o código do desenvolvimento do Modelo a novos dados periodicamente, ou seja, a novos clientes e a clientes cujas características se alteram.

No Anexo 13 encontra-se todo o Source Code do Projecto em linguagem SAS.

---

24 Anexo 7 - Nó Cluster com o método K-Means e o respectivo perfil de segmentos

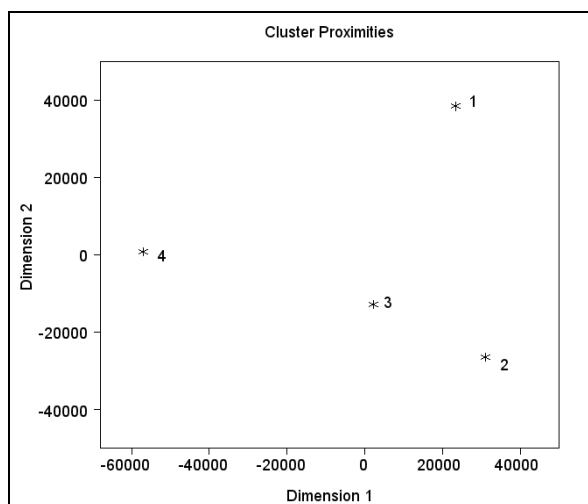
25 As variáveis utilizadas para a análise de perfil foram: Potencial Socioeconómico, Quadrante em que o cliente se insere no mês actual, Potencial Bancário, Desenvolvimento associado a zona habitacional, Negócio preferencial associado a zona habitacional, Nível de riqueza associado a zona habitacional, Composição da Carteira e Escalão de Idade.

## 4. Resultados

### 4.1 Análise dos Segmentos Criados

Como resultado final do desenvolvimento do Modelo de Segmentação de Clientes com posse de produtos de investimento obteve-se 4 segmentos distintos. A distância entre os segmentos encontra-se representada na figura 3. Pode-se então concluir que a solução obtida ( $k=4$ ) revela dissimilaridades nas características ou comportamentos dos clientes que a eles pertençam, considerando-se assim uma boa solução.

Figura 3. Distância entre Clusters



É natural que periodicamente se verifique migração entre os segmentos, uma vez que as características dos clientes são dinâmicas. Assim, conclui-se que poderá existir, de futuro, migração entre os segmentos 2 e 3 (visto serem aqueles que estão mais próximos).

As variáveis mais relevantes para a obtenção desta solução ótima foram o saldo que o cliente tem: (1) em produtos de alto risco nos últimos 12 meses; (2) em produtos de alto risco nos últimos 3 meses; (3) em produtos fiscais nos últimos 6 meses e (4) em produtos de baixo risco no mês actual. Para uma maior compreensão visualizar o Anexo 8.

Após a identificação do número ótimo dos segmentos ( $k=4$ ) é fundamental proceder à análise das características dos mesmos (anexo 9 e anexo 10).

O **segmento 1** é formado, essencialmente, por um elevado número de clientes possuidores de PAR, com especial destaque para a posse de Produtos Estruturados e de Seguros de Capitalização (quase um terço dos clientes têm estes seguros). Assim, este grupo de **clientes tem tendência para o risco** e representa 2,72% da população.

Já o **segmento 2** é composto essencialmente por clientes que não possuem qualquer tipo de produtos de investimento acima dos 250 euros (produtos residuais), possuindo assim uma **carteira diversificada**. Este segmento representa 92,01% da população, o que pode ser explicado dado o perfil deste segmento - perfil do cliente bancário comum.

O **Segmento 3** é caracterizado por clientes que possuem tipicamente PBR (1,88% da população). Este comportamento é ainda mais visível na posse de produtos fiscais (produto considerado de BR). Como tal, este segmento apresenta uma **tendência contrária ao risco**.

No **Segmento 4**, tal como o segmento 3, é caracterizado por 3,4% dos clientes, clientes estes com **maior posse de produtos de baixo risco**, em especial com posse de depósitos a prazo (87% dos clientes).

Após uma identificação e análise dos segmentos foi fundamental analisar o perfil dos mesmos, de acordo com outras variáveis e identificar se o comportamento de cada segmento difere do comportamento da população (nó de Segment Profile). A tabela 1 reflecte as principais características da análise de perfil de cada um dos segmentos.

Tabela 1. Resumo dos Perfis de cada Segmento

Características	Segmento			
	1	2	3	4
Potencial socioeconómico	Médio	Baixo/Médio	Médio/Desconhecido	Médio/Elevado
Quadrante Vínculo Bancário	Alto	Baixo/Médio	Médio/alto	Médio/alto
Potencial Bancário	Baixo	Baixo/Médio	Baixo/Médio	Baixo/Médio
Desenvolvimento do Concelho	Metropolitano/Urbano	Metropolitano/Urbano	Metropolitano/Urbano	Metropolitano/Urbano
Negócio do Concelho	Com Recursos	Com Recursos/Alguns Créditos	Com Recursos/Alguns Créditos	Com Recursos/Alguns Créditos
Nível Riqueza do Concelho	Médio	Médio/Médio Alto	Médio/Médio Alto	Médio/Rico
Escala de Idade	acima dos 40 anos	acima dos 40 anos	acima dos 40 anos	acima dos 40 anos
Portfólio Produtos	FUNDOS*** DPRAZO** PESTRUTURADOS** PFISCAIS** SEGUROS**	DPRAZO*	DPRAZO* FUNDOS* PFISCAIS*** PESTRUTURADOS* SEGUROS*	DPRAZO*** FUNDOS* PESTRUTURADOS** SEGUROS**

As variáveis mais relevantes para explicar cada um dos segmentos encontra-se no anexo 11.

Conclui-se ainda que, tipicamente, as características dos segmentos diferem das características da população, o que significa que a solução encontrada é relevante e explicativa (anexo 12).

## 4.2 Oportunidades/Campanhas sobre cada um dos Segmentos

Perante os diversos perfis de segmentos resultantes deste estudo, apresenta-se na tabela 2 campanhas-alvo para atrair clientes e reter outros a cada um dos Segmentos.

Tabela 2. Oportunidade para cada um dos Segmentos

Segmentos	Oportunidades
1- Agressivos	-Envio de sms/postal de aniversário e no mesmo momento apresenta-se novos produtos de investimento -Reter estes clientes estimulando-os com ofertas(experiências, brindes)
2-Poupados	Para novos clientes: Envio de cartas/SMS Boas Vindas e no mesmo momento apresenta-se produtos de capital garantido (PBR). Para clientes existentes: Contacto directo (telefónico) por parte do gestor de conta a oferecer produtos de capital garantido (PBR) e regalias associadas.
3-Conservadores	Envio de cartas/SMS congratulando-o da antiguidade como cliente e na mesma oportunidade incentiva-se a adquirir mais PBR.
4-Cautelosos	Direccionar cartas de incentivo ao PAR, oferecendo vantagens/regalias específicas a clientes com posse PBR.

## 5. Conclusões e Trabalho Futuro

### 5.1 Conclusões

Perante uma agitação, dinâmica e competitividade dos mercados financeiros é fundamental uma análise do perfil dos clientes com posse de produtos de investimento, de forma a conhecer as características destes e efectuar uma gestão de relacionamento mais eficaz e direccionada.

Após todo o desenvolvimento e implementação do modelo de segmentação de clientes com posse de produtos de investimento, chegou-se à conclusão que os Clientes bancários apresentam um perfil de segmentos resultantes da formação dos clusters diferenciados pelas variáveis que os caracterizam, obtendo como resultado final 4 segmentos distintos. O **Segmento 1** poderá ser denominado de “**Agressivo**” pois é composto maioritariamente por clientes que dado à natureza de produtos que possuem, pretendem investimentos a longo prazo e crescimento de capital, amantes do risco. O **Segmento 2** pode ser visto como o segmento cujos clientes são “**poupados**”, uma vez que é composto predominante por clientes que não possuem qualquer tipo de produtos de investimento, avessos ao risco. O **Segmento 3** é um segmento caracterizado por clientes ponderados nas suas escolhas de produtos que desejam capital garantido sem riscos (**clientes conservadores**). Por fim, temos o **Segmento 4** que é formado por clientes com posse de produtos de investimento de crescimento estável (**clientes cautelosos**). Desta forma obteve-se um conhecimento profundo das características sociodemográficas e de envolvimento dos clientes com posse de produtos de investimento e o enquadramento das mesmas no perfil de cada segmento.

Os segmentos obtidos têm características distintas da população ou mesmo entre si, o que significa que podemos considerar esta solução como óptima.

## 5.2 Trabalho Futuro

No processo de desenvolvimento do modelo de segmentação sugere-se, de futuro, as seguintes melhorias:

- Inclusão de variáveis que retratem o impacto das interacções entre o banco e os clientes;
- Testar a possibilidade de interacções entre variáveis.

Relativamente à aplicação deste modelo sugere-se, como trabalho futuro, a realização das seguintes tarefas:

- Avaliação das migrações existentes entre os segmentos criados, utilizando para tal um novo mês de processamento;
- Realização de campanhas de marketing diferenciadas de acordo com o valor/segmento dos clientes. Desta forma consegue-se reduzir o número de campanhas efectuadas obtendo o melhor rácio custo/benefício e dando enfoque aos clientes “VIP” ou clientes de maior valor já que destes teremos mais retorno;
- Analisar o impacto da tomada de decisão pois é necessário perceber se as acções de marketing diferenciadas sobre os clientes sortiram o efeito desejado (maior retorno). E entender se este impacto teve um balanço positivo ou negativo.



## Bibliografia

- Adriaans, Z. (1996). Data Mining. England: Addison-Wesley.
- Armstrong/Kotler, f. (2003). Marketing: An Introduction . Australia: Pearson International .
- Berry, L. (1983). Relationship Marketing. Chicago, II, 25-8: American Marketing Association.
- Buffet, W. (09 de 2007). "Risk comes from not knowing what you're doing." <http://www.bigfatpurse.com/2007/09/warren-buffett-risk-comes-from-not-knowing-what-youre-doing/> acesso web 13/06/2011.
- C. Laudon, K. (2000). Management Information Systems: Organization and Technology in the Networked Enterprise: 6th Edition. Englewood Cliffs, New Jersey: Prentice Hall.
- Dyche, J. (2003). Improving Business Performance with Analytical CRM. [http://crmguru.custhelp.com/app/answers/detail/a\\_id/872](http://crmguru.custhelp.com/app/answers/detail/a_id/872) acesso 13/06/2011.
- Gorunescu, F. (2011). Data Mining: Concepts, Models and Techniques. Springer-Verlag Berlin Heidelberg.
- Kotler, P. (1998). Administração de Marketing: análise planejamento, implementação e controle. São Paulo: Atlas.
- Kumar, V. P.-N. (2005). Introduction to Data Mining. Addison-Wesley.
- Macedo, J. (2010). Como Classificar um Investimento. <http://www.produtosbancarios.com/como-classificar-um-investimento/> acesso web 13/06/2011.
- Madnick, D. (1977). Institutional and Ad Hoc DSS and Their Effective Use. Santa Clara, California: ACM.
- METAGroup. (2000). APPLICATION DELIVERY STRATEGIES. Harbor Drive, PO Box 120061, Stamford: META Group Inc.
- O'Brien, J. A. (2006). Management Information Systems, 7th Edition. Irwin McGraw-Hill.
- Peppers D. e Rogers, M. (1993). The One to One Future: Building Relationships One Customer at a Time. New York: Double Day.
- Porter, M. (1985). Competitive Advantage, Free Press. New York: Free Press.
- Pregibon, D. (8 de Dezembro de 1996). Data Mining, Statistical Computing & Graphics Newsletter.

Sheth, A. e. (2001). Customer Relationship Management. Customer Relationship Management: Emerging Practice, Process, and Discipline. Journal of Economic and Social Research 3.

Sokal, S. &. (1973). Numerical Taxonomy. San Francisco, CA.: Freeman.

Turban, E. a. (1998). Decision Support Systems and Intelligent Systems. Upper Saddle River, NJ: Prentice Hall.

Woodcock, M. S. (2002). Marketing de Relacionamento. Littera Mundi.

## Anexos

### Anexo 1. Variáveis que constam na ABT de desenvolvimento

Name	Role	Level	Report	Order	Drop
CCLIENTE	ID	Nominal	No		No
CCOD_DISTrito	Input	Nominal	No		No
CCONCELHO_DESENVOL	Input	Ordinal	No		No
CCONCELHO_NEGOCIO	Input	Ordinal	No		No
CCONCELHO_RIQUEZA	Input	Ordinal	No		No
CSATISFACAO_BALCAO	Input	Ordinal	No		No
CSATISFACAO_BANCO	Input	Ordinal	No		No
CSATISFACAO_CANALIS	Input	Ordinal	No		No
ICLIENTE_ACTIVo_MES12	Input	Binary	No		No
ICOLABORADOR_FLG	Input	Binary	No		No
ID_PRAZO	Input	Binary	No		No
IEMIGRANTE_FLG	Input	Binary	No		No
IESCALAO_ANTIGMES12	Input	Binary	No		No
IESCALAO_IDADEMES12	Input	Binary	No		No
IFALECIDO_FLG	Input	Binary	No		No
IFUNDOS	Input	Binary	No		No
INACIONALIDADE_FLG	Input	Binary	No		No
IPOUPANCAS	Input	Binary	No		No
IP_ESTRUT	Input	Binary	No		No
IP_FISCAIS	Input	Binary	No		No
IQUADMES12	Input	Binary	No		No
ISEXO_CD	Input	Binary	No		No
IS_CAPIT	Input	Binary	No		No
POTENCIAL_BANCARIO	Input	Nominal	No		No
POTENCIAL_SE	Input	Nominal	No		No
VMONSALDO_PROD_AR_MES12	Input	Interval	No		No
VMONSALDO_PROD_AR_ULT12M	Input	Interval	No		No
VMONSALDO_PROD_AR_ULT3M	Input	Interval	No		No
VMONSALDO_PROD_AR_ULT6M	Input	Interval	No		No
VMONSALDO_PROD_BR_MES12	Input	Interval	No		No
VMONSALDO_PROD_BR_ULT12M	Input	Interval	No		No
VMONSALDO_PROD_BR_ULT3M	Input	Interval	No		No
VMONSALDO_PROD_BR_ULT6M	Input	Interval	No		No
VMON_MAX_SALDO_DP_ULT12M	Input	Interval	No		No
VMON_MAX_SALDO_FU_ULT12M	Input	Interval	No		No
VMON_MAX_SALDO_PE_ULT12M	Input	Interval	No		No
VMON_MAX_SALDO_PFI_ULT12M	Input	Interval	No		No
VMON_MAX_SALDO_PO_ULT12M	Input	Interval	No		No
VMON_MAX_SALDO_SE_ULT12M	Input	Interval	No		No
VMON_SALDO_DPMES12	Input	Interval	No		No
VMON_SALDO_DP_ULT12M	Input	Interval	No		No
VMON_SALDO_DP_ULT3M	Input	Interval	No		No
VMON_SALDO_DP_ULT6M	Input	Interval	No		No
VMON_SALDO_FUMES12	Input	Interval	No		No
VMON_SALDO_FU_ULT12M	Input	Interval	No		No
VMON_SALDO_FU_ULT3M	Input	Interval	No		No
VMON_SALDO_FU_ULT6M	Input	Interval	No		No
VMON_SALDO_PEMES12	Input	Interval	No		No
VMON_SALDO_PE_ULT12M	Input	Interval	No		No
VMON_SALDO_PE_ULT3M	Input	Interval	No		No
VMON_SALDO_PE_ULT6M	Input	Interval	No		No
VMON_SALDO_PFIMES12	Input	Interval	No		No
VMON_SALDO_PFI_ULT12M	Input	Interval	No		No
VMON_SALDO_PFI_ULT3M	Input	Interval	No		No
VMON_SALDO_PFI_ULT6M	Input	Interval	No		No
VMON_SALDO_POMES12	Input	Interval	No		No
VMON_SALDO_PO_ULT12M	Input	Interval	No		No
VMON_SALDO_PO_ULT3M	Input	Interval	No		No
VMON_SALDO_PO_ULT6M	Input	Interval	No		No
VMON_SALDO_SEMES12	Input	Interval	No		No
VMON_SALDO_SE_ULT12M	Input	Interval	No		No
VMON_SALDO_SE_ULT3M	Input	Interval	No		No
VMON_SALDO_SE_ULT6M	Input	Interval	No		No

## Anexo 2. Código SAS- Pré-tratamento de Missings

```

/*Pré-Tratamento de Missings*/

if ISALDO_DO_FLGMES12 = . THEN ISALDO_DO_FLGMES12=0;

IF CONCELHO_RIQUEZA= . THEN CONCELHO_RIQUEZA=5;

IF CONCELHO_NEGOCIO= . THEN CONCELHO_NEGOCIO=5;


IF CONCELHO_DESENVOL=. THEN CONCELHO_DESENVOL=6;

if POTENCIAL_SE=. then POTENCIAL_SE=5;

```

## Anexo 3. Nó Impute e as respectivas propriedades associadas


Class Variables	
Default Input Method	Default Constant Value
Default Target Method	None
Normalize Values	Yes
Interval Variables	
Default Input Method	Default Constant Value
Default Target Method	None
Default Constant Value	
Default Character Value	D
Default Number Value	0.0

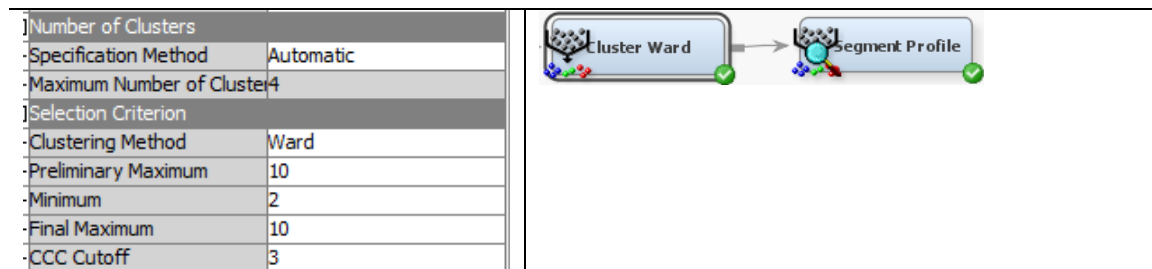
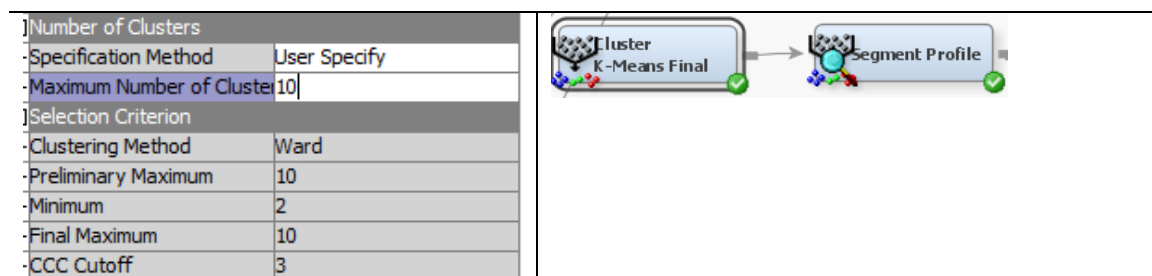
 Impute

## Anexo 4. Nó Filter e as respectivas propriedades associadas

Interval Variables		
Interval Variables		
Default Filtering Method	Extreme Percentiles	
Keep Missing Values	Yes	
Tuning Parameters		

## Anexo 5. Nó Transform Variables com aplicação do método Bucket às variáveis intervalares

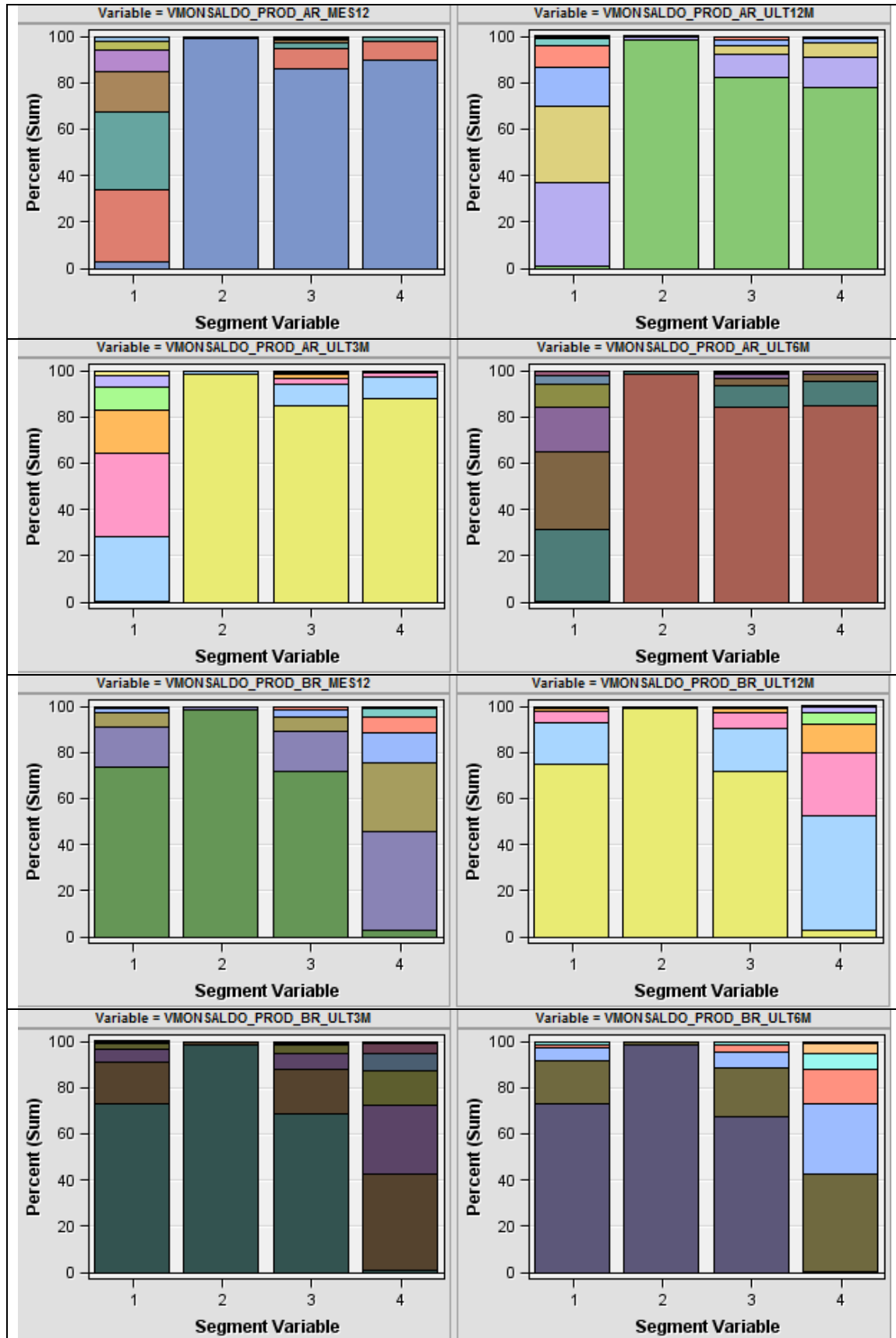
Name	Method	Number of Bins	Role	Level	
IP_ESTRUT	Default	4	Input	Binary	
IP_FISCAIS	Default	4	Input	Binary	
IQUADMES12	Default	4	Input	Binary	
ISALDO_DO_FL	Default	4	Input	Binary	
ISEXO_CD	Default	4	Input	Binary	
IS_CAPIT	Default	4	Input	Binary	
POTENCIAL_SE	Default	4	Input	Nominal	
VMONSALDO_PP	Bucket	4	Input	Interval	
VMONSALDO_PP	Bucket	4	Input	Interval	

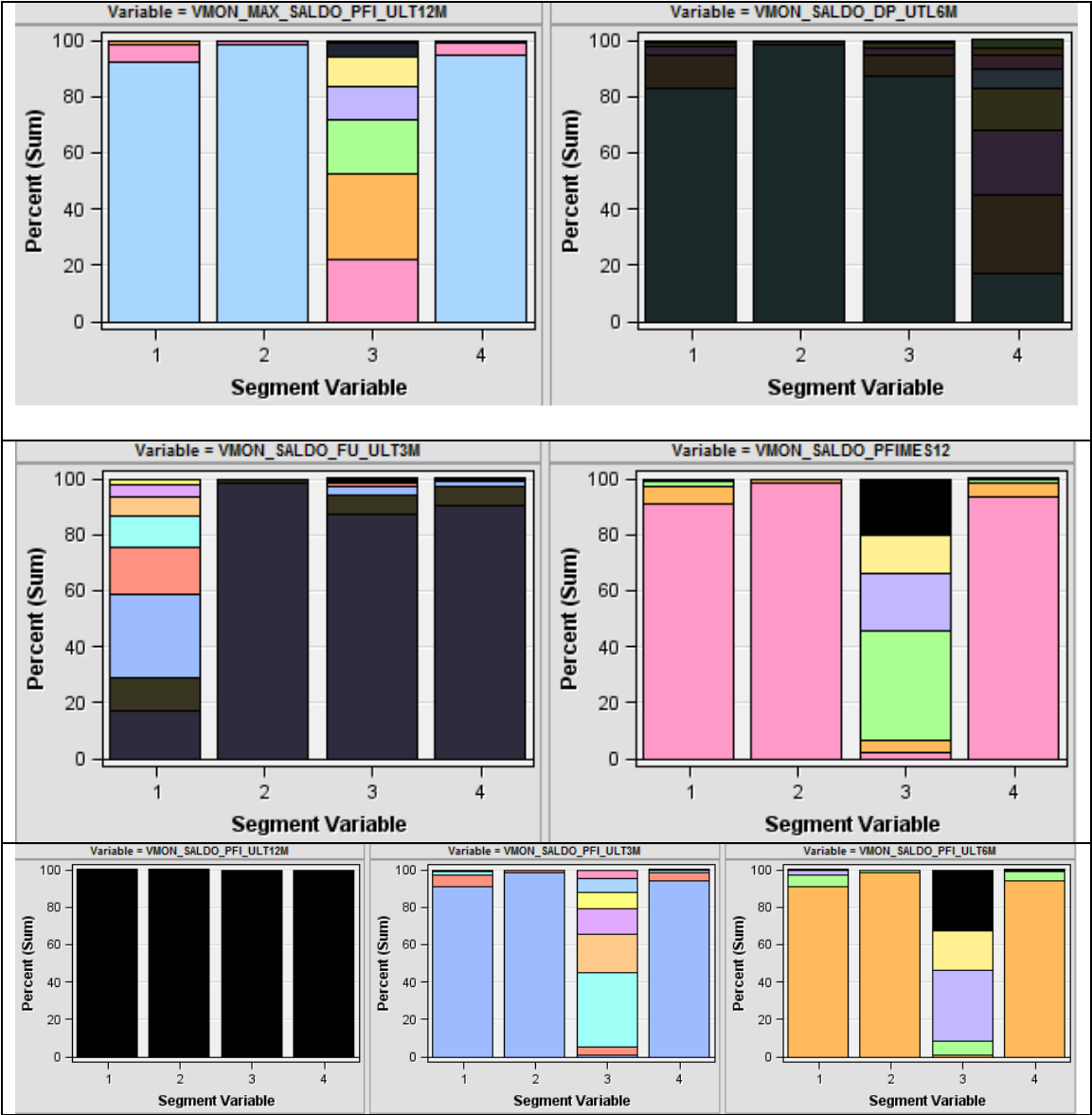
**Anexo 6.** Nó Cluster com o método Ward e o respectivo perfil de segmentos**Anexo 7.** Nó Cluster com o método K-Means com a definição do número de clusters a ser construído e o respectivo perfil de segmentos**Anexo 8.** Variáveis mais relevantes para a solução dos Segmentos obtidos

Família	Varável	Descrição	Importância
PAR	VMONSALDO_PROD_AR_ULT12M	Saldo que o cliente possui em PAR nos últimos 12 meses	1
PAR	VMONSALDO_PROD_AR_ULT3M	Saldo que o cliente possui em Produtos AR nos últimos 3 meses	0,91
Carteira de Produtos	VMON_SALDO_PFI_ULT6M	Saldo da conta de produtos fiscais que o cliente possui nos últimos 6 meses	0,88
PBR	VMONSALDO_PROD_BR_MES12	Saldo que o cliente possui em PBR no mês actual	0,77
PBR	VMONSALDO_PROD_BR_ULT3M	Saldo que o cliente possui em PBR nos últimos 3 meses	0,77
PBR	VMONSALDO_PROD_BR_ULT6M	Saldo que o cliente possui em PBR nos últimos 6 meses	0,77
PBR	VMONSALDO_PROD_BR_ULT12M	Saldo que o cliente possui em PBR nos últimos 12 meses	0,76
Carteira de Produtos	VMON_SALDO_DP_UTL6M	Saldo da conta de depósitos a prazo que o cliente possui nos últimos 3 meses	0,74
PAR	VMONSALDO_PROD_AR_MES12	Saldo que o cliente possui em PAR no mês actual	0,73
PAR	VMONSALDO_PROD_AR_ULT6M	Saldo que o cliente possui em PAR nos últimos 6 meses	0,72
Carteira de Produtos	VMON_SALDO_FU_ULT3M	Saldo da conta de fundos que o cliente possui nos últimos 3 meses	0,69
Carteira de Produtos	VMON_SALDO_PFI_ULT12M	Saldo da conta de produtos fiscais que o cliente possui nos últimos 12 meses	0,59
Carteira de Produtos	VMON_MAX_SALDO_PFI_ULT12M	Saldo máximo da conta de produtos fiscais que o cliente possui nos últimos 12 meses	0,59
Carteira de Produtos	VMON_SALDO_PFI_ULT3M	Saldo da conta de produtos fiscais que o cliente possui nos últimos 3 meses	0,59
Carteira de Produtos	VMON_SALDO_PFIMES12	Saldo da conta de produtos fiscais que o cliente possui no mês actual	0,58

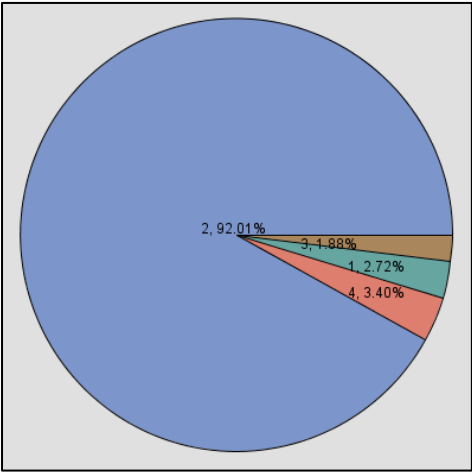
(PAR) – Produtos Alto Risco  
(PBR) -Produtos Baixo Risco

### Anexo 9. Caracterização dos Segmentos





Anexo 10. Dimensão dos Segmentos

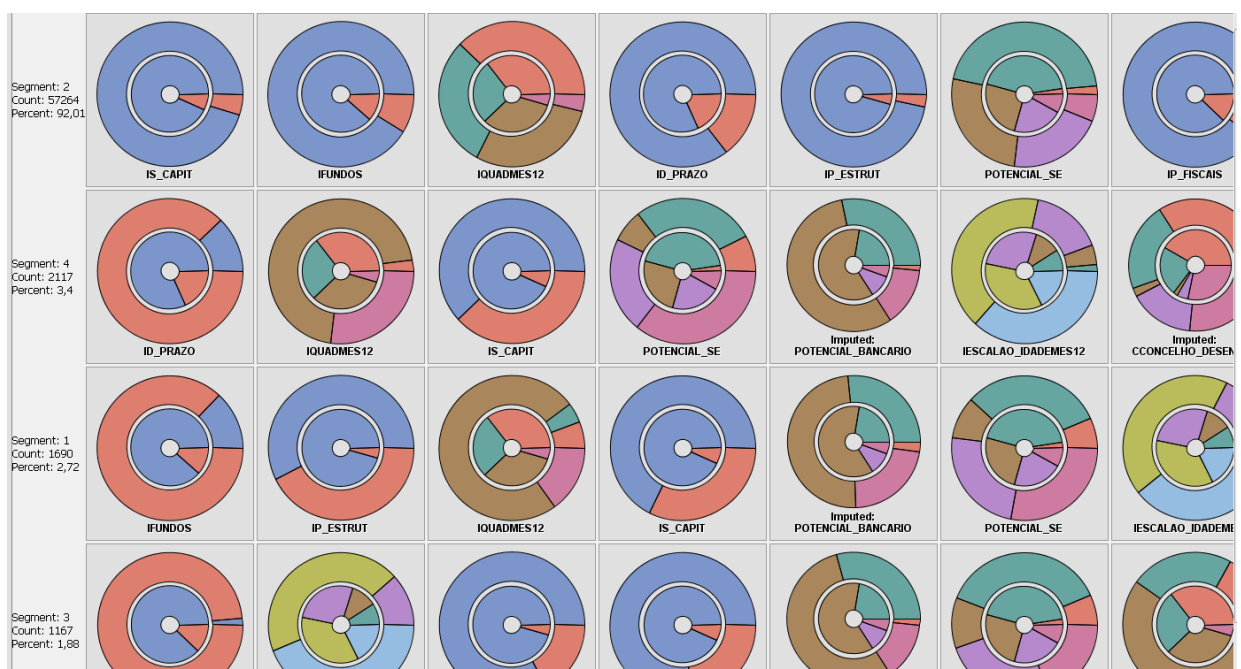


### Anexo 11. Variáveis mais relevantes para cada um dos Segmentos obtidos

Segmento				Classificação
1	2	3	4	
IFUNDOS	IS_CAPIT	IP_FISCAIS	ID_PRAZO	1
IP_ESTRUT	IFUNDOS	IESCALAO_IDADEMES12	IQUADMES12	2
IQUADMES12	IQUADMES12	IP_ESTRUT	IS_CAPIT	3
IS_CAPIT	ID_PRAZO	IS_CAPIT	POTENCIAL_SE	4
POTENCIAL_BANCARIO	IP_ESTRUT	POTENCIAL_BANCARIO	POTENCIAL_BANCARIO	5
POTENCIAL_SE	POTENCIAL_SE	POTENCIAL_SE	IESCALAO_IDADEMES12	6
IESCALAO_IDADEMES12	IP_FISCAIS	IQUADMES12	CCONCELHO_DESENVOL	7
ID_PRAZO	POTENCIAL_BANCARIO	IPOUPANCAS	IP_ESTRUT	8
CCONCELHO_RIQUEZA	IESCALAO_IDADEMES12	IFUNDOS	CCONCELHO_RIQUEZA	9
CCONCELHO_DESENVOL	CCONCELHO_RIQUEZA	ID_PRAZO	CCONCELHO_NEGOCIO	10
CCONCELHO_NEGOCIO	CCONCELHO_DESENVOL	CCONCELHO_DESENVOL	IFUNDOS	*
IPOUPANCAS	CCONCELHO_NEGOCIO	CCONCELHO_NEGOCIO	IPOUPANCAS	*
IP_FISCAIS	IPOUPANCAS	CCONCELHO_RIQUEZA	IP_FISCAIS	*

\*-Não contribuem para o peso da relevância no respectivo segmento

## **Anexo 12.** Comparação das características dos Segmento face às características da População







## Anexo 13. Source Code do Projecto

```

*-----*;
* EM SCORE CODE;
* VERSION: 6.2;
* GENERATED BY: Marcelo;
* CREATED: 23JUL2011:00:33:36;
*-----*;
*-----*;
* TOOL: Input Data Source;
* TYPE: SAMPLE;
* NODE: Ids;
*-----*;
*-----*;
* TOOL: SASHELP.EMCORE.EMCODETOOL.CLASS;
* TYPE: UTILITY;
* NODE: ECODE;
*-----*;
*-----*;
* TOOL: Metadata Node;
* TYPE: UTILITY;
* NODE: Meta;
*-----*;
*-----*;
* TOOL: Imputation;
* TYPE: MODIFY;
* NODE: Impt;
*-----*;
*-----*;
* TOOL: Filtering;
* TYPE: MODIFY;
* NODE: Filter;
*-----*;
*-----*;
* TOOL: Clustering;
* TYPE: EXPLORE;
* NODE: Clus3;
*-----*;
*****;
*** Begin Scoring Code from PROC DMVQ ***;
*****;

*** Begin Class Look-up, Standardization, Replacement ;

```

```
drop _dm_bad; _dm_bad = 0;
```

```
*** Standardize VMONSALDO_PROD_AR_MES12 ;  
drop T_VMONSALDO_PROD_AR_MES12 ;  
if missing( VMONSALDO_PROD_AR_MES12 ) then T_VMONSALDO_PROD_AR_MES12 = .;  
else T_VMONSALDO_PROD_AR_MES12 = (VMONSALDO_PROD_AR_MES12  
    - 0) * 0.00001192972058;
```

```
*** Standardize VMONSALDO_PROD_AR_ULT12M ;  
drop T_VMONSALDO_PROD_AR_ULT12M ;  
if missing( VMONSALDO_PROD_AR_ULT12M ) then T_VMONSALDO_PROD_AR_ULT12M = .;  
else T_VMONSALDO_PROD_AR_ULT12M = (VMONSALDO_PROD_AR_ULT12M  
    - 0) * 0.00001045621894;
```

```
*** Standardize VMONSALDO_PROD_AR_ULT3M ;  
drop T_VMONSALDO_PROD_AR_ULT3M ;  
if missing( VMONSALDO_PROD_AR_ULT3M ) then T_VMONSALDO_PROD_AR_ULT3M = .;  
else T_VMONSALDO_PROD_AR_ULT3M = (VMONSALDO_PROD_AR_ULT3M  
    - 0) * 0.00001222881083;
```

```
*** Standardize VMONSALDO_PROD_AR_ULT6M ;  
drop T_VMONSALDO_PROD_AR_ULT6M ;  
if missing( VMONSALDO_PROD_AR_ULT6M ) then T_VMONSALDO_PROD_AR_ULT6M = .;  
else T_VMONSALDO_PROD_AR_ULT6M = (VMONSALDO_PROD_AR_ULT6M  
    - 0) * 0.00001156379715;
```

```
*** Standardize VMONSALDO_PROD_BR_MES12 ;  
drop T_VMONSALDO_PROD_BR_MES12 ;  
if missing( VMONSALDO_PROD_BR_MES12 ) then T_VMONSALDO_PROD_BR_MES12 = .;  
else T_VMONSALDO_PROD_BR_MES12 = (VMONSALDO_PROD_BR_MES12  
    - 0) * 6.7796610169491E-6;
```

```
*** Standardize VMONSALDO_PROD_BR_ULT12M ;  
drop T_VMONSALDO_PROD_BR_ULT12M ;  
if missing( VMONSALDO_PROD_BR_ULT12M ) then T_VMONSALDO_PROD_BR_ULT12M = .;  
else T_VMONSALDO_PROD_BR_ULT12M = (VMONSALDO_PROD_BR_ULT12M  
    - 0) * 6.2741757098646E-6;
```

```
*** Standardize VMONSALDO_PROD_BR_ULT3M ;  
drop T_VMONSALDO_PROD_BR_ULT3M ;  
if missing( VMONSALDO_PROD_BR_ULT3M ) then T_VMONSALDO_PROD_BR_ULT3M = .;  
else T_VMONSALDO_PROD_BR_ULT3M = (VMONSALDO_PROD_BR_ULT3M  
    - 0) * 7.1428571428571E-6;
```

```
*** Standardize VMONSALDO_PROD_BR_ULT6M ;
```

```
drop T_VMONSALDO_PROD_BR_ULT6M ;
if missing( VMONSALDO_PROD_BR_ULT6M ) then T_VMONSALDO_PROD_BR_ULT6M = .;
else T_VMONSALDO_PROD_BR_ULT6M = (VMONSALDO_PROD_BR_ULT6M
- 0) * 7.1356450444729E-6;

*** Standardize VMON_MAX_SALDO_PFI_ULT12M ;
drop T_VMON_MAX_SALDO_PFI_ULT12M ;
if missing( VMON_MAX_SALDO_PFI_ULT12M ) then T_VMON_MAX_SALDO_PFI_ULT12M = .;
else T_VMON_MAX_SALDO_PFI_ULT12M = (VMON_MAX_SALDO_PFI_ULT12M
- 0) * 0.00004148785401;

*** Standardize VMON_SALDO_DP_UTL6M ;
drop T_VMON_SALDO_DP_UTL6M ;
if missing( VMON_SALDO_DP_UTL6M ) then T_VMON_SALDO_DP_UTL6M = .;
else T_VMON_SALDO_DP_UTL6M = (VMON_SALDO_DP_UTL6M - 0) * 9.4969341047629E-6;

*** Standardize VMON_SALDO_FU_ULT3M ;
drop T_VMON_SALDO_FU_ULT3M ;
if missing( VMON_SALDO_FU_ULT3M ) then T_VMON_SALDO_FU_ULT3M = .;
else T_VMON_SALDO_FU_ULT3M = (VMON_SALDO_FU_ULT3M - 0) * 0.00001694308175;

*** Standardize VMON_SALDO_PFIMES12 ;
drop T_VMON_SALDO_PFIMES12 ;
if missing( VMON_SALDO_PFIMES12 ) then T_VMON_SALDO_PFIMES12 = .;
else T_VMON_SALDO_PFIMES12 = (VMON_SALDO_PFIMES12 - 0) * 0.0000496162924;

*** Standardize VMON_SALDO_PFI_ULT12M ;
drop T_VMON_SALDO_PFI_ULT12M ;
if missing( VMON_SALDO_PFI_ULT12M ) then T_VMON_SALDO_PFI_ULT12M = .;
else T_VMON_SALDO_PFI_ULT12M = (VMON_SALDO_PFI_ULT12M - 0) * 0.00004584221605;

*** Standardize VMON_SALDO_PFI_ULT3M ;
drop T_VMON_SALDO_PFI_ULT3M ;
if missing( VMON_SALDO_PFI_ULT3M ) then T_VMON_SALDO_PFI_ULT3M = .;
else T_VMON_SALDO_PFI_ULT3M = (VMON_SALDO_PFI_ULT3M - 0) * 0.00004983221493;

*** Standardize VMON_SALDO_PFI_ULT6M ;
drop T_VMON_SALDO_PFI_ULT6M ;
if missing( VMON_SALDO_PFI_ULT6M ) then T_VMON_SALDO_PFI_ULT6M = .;
else T_VMON_SALDO_PFI_ULT6M = (VMON_SALDO_PFI_ULT6M - 0) * 0.0000483449392;

*** End Class Look-up, Standardization, Replacement ;

*** Omitted Cases;
```

```

if _dm_bad then do;
  _SEGMENT_ = .; Distance = .;
  goto CLUS3vlex ;
end; *** omitted;

*** Compute Distances and Cluster Membership;
label _SEGMENT_ = 'Segment Id' ;
label Distance = 'Distance' ;
array CLUS3vads [4] _temporary_;
drop _vqclus _vqmvar _vqnvar;
_vqmvar = 0;
do _vqclus = 1 to 4; CLUS3vads [_vqclus] = 0; end;
if not missing( T_VMONSALDO_PROD_AR_MES12 ) then do;
  CLUS3vads [1] + ( T_VMONSALDO_PROD_AR_MES12 - 0.34030018152168 )**2;
  CLUS3vads [2] + ( T_VMONSALDO_PROD_AR_MES12 - 0.0062413897022 )**2;
  CLUS3vads [3] + ( T_VMONSALDO_PROD_AR_MES12 - 0.05016262666719 )**2;
  CLUS3vads [4] + ( T_VMONSALDO_PROD_AR_MES12 - 0.03437511521451 )**2;
end;
else _vqmvar + 0.00457333699611;
if not missing( T_VMONSALDO_PROD_AR_ULT12M ) then do;
  CLUS3vads [1] + ( T_VMONSALDO_PROD_AR_ULT12M - 0.32808422795335 )**2;
  CLUS3vads [2] + ( T_VMONSALDO_PROD_AR_ULT12M - 0.00779956676402 )**2;
  CLUS3vads [3] + ( T_VMONSALDO_PROD_AR_ULT12M - 0.06273647278404 )**2;
  CLUS3vads [4] + ( T_VMONSALDO_PROD_AR_ULT12M - 0.07087565667216 )**2;
end;
else _vqmvar + 0.00482179387042;
if not missing( T_VMONSALDO_PROD_AR_ULT3M ) then do;
  CLUS3vads [1] + ( T_VMONSALDO_PROD_AR_ULT3M - 0.35979263758268 )**2;
  CLUS3vads [2] + ( T_VMONSALDO_PROD_AR_ULT3M - 0.00667729567626 )**2;
  CLUS3vads [3] + ( T_VMONSALDO_PROD_AR_ULT3M - 0.05450998263873 )**2;
  CLUS3vads [4] + ( T_VMONSALDO_PROD_AR_ULT3M - 0.03880078106274 )**2;
end;
else _vqmvar + 0.00499182164873;
if not missing( T_VMONSALDO_PROD_AR_ULT6M ) then do;
  CLUS3vads [1] + ( T_VMONSALDO_PROD_AR_ULT6M - 0.35062179978964 )**2;
  CLUS3vads [2] + ( T_VMONSALDO_PROD_AR_ULT6M - 0.00706971288035 )**2;
  CLUS3vads [3] + ( T_VMONSALDO_PROD_AR_ULT6M - 0.05760423593536 )**2;
  CLUS3vads [4] + ( T_VMONSALDO_PROD_AR_ULT6M - 0.04883811674173 )**2;
end;
else _vqmvar + 0.0049158253547;
if not missing( T_VMONSALDO_PROD_BR_MES12 ) then do;
  CLUS3vads [1] + ( T_VMONSALDO_PROD_BR_MES12 - 0.08373029167087 )**2;
  CLUS3vads [2] + ( T_VMONSALDO_PROD_BR_MES12 - 0.00954617519957 )**2;
  CLUS3vads [3] + ( T_VMONSALDO_PROD_BR_MES12 - 0.1193491883013 )**2;
  CLUS3vads [4] + ( T_VMONSALDO_PROD_BR_MES12 - 0.29952235694585 )**2;

```

```

end;
else _vqmvar + 0.00503941934754;
if not missing( T_VMONSALDO_PROD_BR_ULT12M ) then do;
  CLUS3vads [1] + ( T_VMONSALDO_PROD_BR_ULT12M - 0.07890194981277 )**2;
  CLUS3vads [2] + ( T_VMONSALDO_PROD_BR_ULT12M - 0.00972313570284 )**2;
  CLUS3vads [3] + ( T_VMONSALDO_PROD_BR_ULT12M - 0.11358757922142 )**2;
  CLUS3vads [4] + ( T_VMONSALDO_PROD_BR_ULT12M - 0.27486369064032 )**2;
end;
else _vqmvar + 0.00429107792069;
if not missing( T_VMONSALDO_PROD_BR_ULT3M ) then do;
  CLUS3vads [1] + ( T_VMONSALDO_PROD_BR_ULT3M - 0.08654161291787 )**2;
  CLUS3vads [2] + ( T_VMONSALDO_PROD_BR_ULT3M - 0.01012391707828 )**2;
  CLUS3vads [3] + ( T_VMONSALDO_PROD_BR_ULT3M - 0.12594626853099 )**2;
  CLUS3vads [4] + ( T_VMONSALDO_PROD_BR_ULT3M - 0.31502163693511 )**2;
end;
else _vqmvar + 0.00542939679581;
if not missing( T_VMONSALDO_PROD_BR_ULT6M ) then do;
  CLUS3vads [1] + ( T_VMONSALDO_PROD_BR_ULT6M - 0.08746763894678 )**2;
  CLUS3vads [2] + ( T_VMONSALDO_PROD_BR_ULT6M - 0.01039271972571 )**2;
  CLUS3vads [3] + ( T_VMONSALDO_PROD_BR_ULT6M - 0.12651553450427 )**2;
  CLUS3vads [4] + ( T_VMONSALDO_PROD_BR_ULT6M - 0.3150016785505 )**2;
end;
else _vqmvar + 0.00541307738257;
if not missing( T_VMON_MAX_SALDO_PFI_ULT12M ) then do;
  CLUS3vads [1] + ( T_VMON_MAX_SALDO_PFI_ULT12M - 0.0236040981803 )**2;
  CLUS3vads [2] + ( T_VMON_MAX_SALDO_PFI_ULT12M - 0.00655600218855 )**2;
  CLUS3vads [3] + ( T_VMON_MAX_SALDO_PFI_ULT12M - 0.4139466838179 )**2;
  CLUS3vads [4] + ( T_VMON_MAX_SALDO_PFI_ULT12M - 0.01691023849163 )**2;
end;
else _vqmvar + 0.0044445452614;
if not missing( T_VMON_SALDO_DP_UTL6M ) then do;
  CLUS3vads [1] + ( T_VMON_SALDO_DP_UTL6M - 0.05839232063328 )**2;
  CLUS3vads [2] + ( T_VMON_SALDO_DP_UTL6M - 0.00829914632512 )**2;
  CLUS3vads [3] + ( T_VMON_SALDO_DP_UTL6M - 0.04512449087727 )**2;
  CLUS3vads [4] + ( T_VMON_SALDO_DP_UTL6M - 0.31257402095397 )**2;
end;
else _vqmvar + 0.00598582614181;
if not missing( T_VMON_SALDO_FU_ULT3M ) then do;
  CLUS3vads [1] + ( T_VMON_SALDO_FU_ULT3M - 0.35915418773254 )**2;
  CLUS3vads [2] + ( T_VMON_SALDO_FU_ULT3M - 0.00626781766165 )**2;
  CLUS3vads [3] + ( T_VMON_SALDO_FU_ULT3M - 0.0451083647279 )**2;
  CLUS3vads [4] + ( T_VMON_SALDO_FU_ULT3M - 0.02948292333279 )**2;
end;
else _vqmvar + 0.00582341199014;
if not missing( T_VMON_SALDO_PFIMES12 ) then do;

```

```

    CLUS3vads [1] + ( T_VMON_SALDO_PFI_ULT12M - 0.0262243628411 )**2;
    CLUS3vads [2] + ( T_VMON_SALDO_PFI_ULT12M - 0.00705755792665 )**2;
    CLUS3vads [3] + ( T_VMON_SALDO_PFI_ULT12M - 0.45195998742013 )**2;
    CLUS3vads [4] + ( T_VMON_SALDO_PFI_ULT12M - 0.01725835449051 )**2;
end;
else _vqmvar + 0.00530594161669;
if not missing( T_VMON_SALDO_PFI_ULT12M ) then do;
    CLUS3vads [1] + ( T_VMON_SALDO_PFI_ULT12M - 0.02384401673743 )**2;
    CLUS3vads [2] + ( T_VMON_SALDO_PFI_ULT12M - 0.00628785774858 )**2;
    CLUS3vads [3] + ( T_VMON_SALDO_PFI_ULT12M - 0.42771775774882 )**2;
    CLUS3vads [4] + ( T_VMON_SALDO_PFI_ULT12M - 0.01656508162908 )**2;
end;
else _vqmvar + 0.00464500460363;
if not missing( T_VMON_SALDO_PFI_ULT3M ) then do;
    CLUS3vads [1] + ( T_VMON_SALDO_PFI_ULT3M - 0.02609147306343 )**2;
    CLUS3vads [2] + ( T_VMON_SALDO_PFI_ULT3M - 0.00699723165935 )**2;
    CLUS3vads [3] + ( T_VMON_SALDO_PFI_ULT3M - 0.45676461883997 )**2;
    CLUS3vads [4] + ( T_VMON_SALDO_PFI_ULT3M - 0.01730364975329 )**2;
end;
else _vqmvar + 0.00535413055299;
if not missing( T_VMON_SALDO_PFI_ULT6M ) then do;
    CLUS3vads [1] + ( T_VMON_SALDO_PFI_ULT6M - 0.02530414792115 )**2;
    CLUS3vads [2] + ( T_VMON_SALDO_PFI_ULT6M - 0.00671500690536 )**2;
    CLUS3vads [3] + ( T_VMON_SALDO_PFI_ULT6M - 0.44633135290547 )**2;
    CLUS3vads [4] + ( T_VMON_SALDO_PFI_ULT6M - 0.01683245219345 )**2;
end;
else _vqmvar + 0.00508909072949;
_vqnvar = 0.07612370021279 - _vqmvar;
if _vqnvar <= 1.298139412929E-13 then do;
    _SEGMENT_ = .; Distance = .;
end;
else do;
    _SEGMENT_ = 1; Distance = CLUS3vads [1];
    _vqfzdst = Distance * 0.99999999999988; drop _vqfzdst;
    do _vqclus = 2 to 4;
        if CLUS3vads [_vqclus] < _vqfzdst then do;
            _SEGMENT_ = _vqclus; Distance = CLUS3vads [_vqclus];
            _vqfzdst = Distance * 0.99999999999988;
        end;
    end;
    Distance = sqrt(Distance * (0.07612370021279 / _vqnvar));
end;
CLUS3vlex ;;

```

```

*****

```

```

*** End Scoring Code from PROC DMVQ ***;
*****;
*-----*;
* Clus3: Creating Segment Label;
*-----*;
length _SEGMENT_LABEL_ $80;
label _SEGMENT_LABEL_='Segment Description';
if _SEGMENT_ = 1 then _SEGMENT_LABEL_="Cluster1";
else
if _SEGMENT_ = 2 then _SEGMENT_LABEL_="Cluster2";
else
if _SEGMENT_ = 3 then _SEGMENT_LABEL_="Cluster3";
else
if _SEGMENT_ = 4 then _SEGMENT_LABEL_="Cluster4";
*-----*;
* TOOL: Extension Class;
* TYPE: ASSESS;
* NODE: Prof3;
*-----*;
*-----*;
* TOOL: Score Node;
* TYPE: ASSESS;
* NODE: Score;
*-----*;
*-----*;
* Score: Creating Fixed Names;
*-----*;

LABEL EM_SEGMENT = 'Segment Variable';
EM_SEGMENT = _SEGMENT_;

```