

Universidade Lusófona de Humanidades e Tecnologia



**Estudo do perfil de Abandono
dos clientes pertencentes a uma Organização Bancária,
através da aplicação de modelos e técnicas preditivas
de Data Mining**

Luís Miguel Peres Barros Barriga

(20080232)

Relatório de Projecto de Final de Curso

Licenciatura em Informática de Gestão

Orientadora:

Prof. Ana Margarida Gomes Alexandre

Setembro 2011

Índice

Índice de Figuras	5
Índice de Tabelas	6
Resumo	7
Abstract	8
Agradecimentos	9
1. Introdução	10
1.1) Relevância do Modelo de Abandono de Cliente na actualidade	11
2. Enquadramento Teórico	12
2.1) Sistemas de Informação	12
2.2) Sistemas de Suporte à Decisão	13
2.3) Data Warehouse	14
2.4) Gestão da Relação com os Clientes	14
2.5) Data Mining	16
2.5.1) Modelos Analíticos Preditivos	16
2.5.2) Técnicas utilizadas	17
3. Método	18
3.1) Introdução	18
3.2) Metodologia	18
3.3) Descrição dos Dados	21
3.4) Desenvolvimento do Modelo Preditivo	23
3.4.1.) Amostragem e partição	24
3.4.2.) Exploração e transformação	25
3.4.3.) Modelação	26
3.4.4.) Análise do melhor modelo	28
3.4.5.) Implementação do melhor modelo	30
4. Resultados	31

4.1) Identificação das variáveis mais relevantes para explicar o Abandono de Clientes no Banco.	31
4.2) Assessment Score Ranking	31
4.3) Acções de Marketing	33
5. Conclusões e Trabalho Futuro	34
5.1) Conclusões	34
5.2) Trabalho Futuro	35
6. Bibliografia	36
7. Anexos	37
7.1) Anexo 1: Descrição de variáveis da ABT	37
7.2) Anexo 2: Distribuição de valores das variáveis sócio-demográficas	47
7.3) Anexo 3: SAS Code utilizado	50
7.4) Anexo 4: Node Sample	50
7.5) Anexo 5: Node Data Partition	51
7.6) Anexo 6: Node Impute	51
7.7) Anexo 7: Relacionamento da variável Target com as Sócio-Demográficas.	52
7.8) Anexo 8: Node Filter	53
7.9) Anexo 9: Transformation Nodes	53
7.10) Anexo 10: Node Variable Selection	54
7.11) Anexo 11: Ranking de importância das variáveis	54
7.12) Anexo 12: Assessment Score Ranking	55
7.13) Anexo 13: Gráficos de apoio a Assessment Score Ranking	56
7.14) Anexo 14: Extrapolação de Custo/Benefício a partir de tabela de Assessment Score Ranking	57

Índice de Figuras

<i>Figura 1: Metodologia de Plataformas Inteligentes</i>	<u>18</u>
<i>Figura 2 – Horizonte temporal para abandono</i>	<u>19</u>
<i>Figura 4: Diagrama de Desenvolvimento do Modelo Preditivo.</i>	<u>23</u>
<i>Figura 5 – Gráfico Cumulative % Response do modelo vencedor</i>	<u>56</u>
<i>Figura 6 – Gráfico de Lift do modelo vencedor</i>	<u>56</u>

Índice de Tabelas

<i>Tabela 1– Significado dos valores da variável target</i>	<i>21</i>
<i>Tabela 2– Correpondência de Modelos e Transformações</i>	<i>27</i>
<i>Tabela 3 - Designação final dos modelos</i>	<i>28</i>
<i>Tabela 4 - Avaliação dos Modelos</i>	<i>29</i>
<i>Tabela 5 - Event Classification Table (Validation)</i>	<i>30</i>
<i>Tabela 6 - Actuações sugeridas sobre os clientes</i>	<i>33</i>
<i>Tabela 7 - Tipos de variáveis da ABT</i>	<i>37</i>
<i>Tabela 8 – Descrição de variáveis da ABT</i>	<i>46</i>
<i>Tabela 9 – Distribuição de valores da variável target</i>	<i>47</i>
<i>Tabela 10 - Distribuição de clientes por escalão etário</i>	<i>47</i>
<i>Tabela 11 – Distribuição de clientes por género</i>	<i>47</i>
<i>Tabela 12 – Distribuição de clientes por Estado Civil</i>	<i>47</i>
<i>Tabela 13 – Distribuição de clientes por nacionalidade</i>	<i>48</i>
<i>Tabela 14 - Distribuição de clientes por Habilitações Literárias</i>	<i>48</i>
<i>Tabela 15 - Distribuição de clientes por Potencial Bancário</i>	<i>48</i>
<i>Tabela 16 - Distribuição de clientes por Potencial Sócio Económico</i>	<i>48</i>
<i>Tabela 17 - Distribuição de clientes por tipo de concelho (desenvolvimento)</i>	<i>49</i>
<i>Tabela 18 - Distribuição de clientes por tipo de concelho (riqueza)</i>	<i>49</i>
<i>Tabela 19 - Distribuição de clientes por tipo de concelho (negócio)</i>	<i>49</i>
<i>Tabela 20– Ranking de importância das variáveis</i>	<i>55</i>
<i>Tabela 21 – Assessment Score Ranking</i>	<i>55</i>
<i>Tabela 22 – Extrapolação de Custo/Benefício a partir de tabela de Assessment Score Ranking</i>	<i>57</i>

Resumo

O projecto que aqui se apresenta aborda um tema cada vez mais importante, no contexto da actual crise global: retenção (e abandono) dos clientes do mercado Bancário.

Na concretização deste trabalho foi desenvolvido um Modelo Preditivo de Data Mining que decerto ajudará a detectar e a explicar as razões inerentes às situações de Abandono.

A metodologia utilizada no desenvolvimento do projecto foi a metodologia “implementação plataformas inteligentes” (IPI), a qual é utilizada pelo SAS para o desenvolvimento de projectos de Data Mining.

O conhecimento daqui proveniente possibilitará á organização definir estratégias melhor direccionadas para a mitigação antecipada do risco de Abandono conquistando assim uma preciosa vantagem competitiva sobre a concorrência .

Palavras – Chave: Mineralização de Dados, Modelos Preditivos, Abandono, Banca, Gestão do Relacionamento dos Clientes

Abstract

The project here presented addresses an increasingly important issue in the current's global crisis context: retention (or churn) of the Banking market customers.

On completion of this work we have developed a Data Mining's Predictive Data Model which certainly will help detect and explain the reasons behind the churn situations.

The methodology used for developing the project was "implementing intelligent platforms" (IPI), which is used by SAS for the development of Data Mining projects.

The knowledge obtained from here, will empower organizations to better define and target their strategies to mitigate the early churn's risk, thus gaining a valuable competitive advantage over their contenders.

Keywords: Data Mining, Predictive Models, Churn, Banking, Customer Relationship Management

Agradecimentos

Á minha filha Rita, pelos afectos espontâneos que me concedeste, pelos que me tentaste roubar e pelas tantas vezes que conseguiste resgatar a minha atenção para ti própria, durante o tempo deste projecto. Nunca me deixaste esquecer quais são as coisas verdadeiramente importantes da vida.

Ao meu filho Vasco, pela pureza, pela cumplicidade, pelo companheirismo, pelo orgulho. És a materialização da minha necessidade do Belo

Á Célia, pelo carinho, pelo apoio, pela confiança, pelo tempo que me proporcionaste para dedicar a este projecto.

O agradecimento final é inteiramente dedicado à minha Orientadora, Professora Ana Margarida Gomes Alexandre, pela disponibilidade, pelo incentivo, pela inspiração, pela dedicação e companheirismo. Sabe perfeitamente que a considero uma Professora muito competente e uma pessoa muito especial. Bem haja por tudo.

1. Introdução

A crise global que se vive actualmente, reflectida numa vertente de forte retracção da Economia, tem vindo a promover um aumento de instabilidade na tradicionalmente estável relação entre a Banca e os seus clientes, conduzindo a situações de ruptura ou redução de envolvimento entre as partes. Por um lado, os clientes particulares de menores recursos tende a recorrer com maior frequência às suas poupanças para fazer face a situações drásticas de instabilidade, como por exemplo a de desemprego, decrementando ou mesmo anulando o seu valor de cliente. Já os clientes de maiores recursos tende a ser mais selectiva e exigente com a rentabilidade dos seus investimentos, procurando acompanhar e ultrapassar as subidas das taxas de juro, dispondo-se a transferir o seu pecúlio para a concorrência ao mínimo sinal de insatisfação.

Ambos os cenários, extremos e distintos, ilustram situações potenciais de Abandono. É fulcral apostar no incremento do grau de satisfação dos clientes e por inerência na sua retenção, focando esta política, particularmente, nos clientes de maiores recursos.

A satisfação do cliente é uma avaliação pós-consumo em que a alternativa escolhida, no mínimo, alcança ou mesmo excede as expectativas (ROSSI, 1995).

Manter o cliente satisfeito, gerir a sua expectativa de base e superá-la, constitui um aspecto crucial na estratégia comercial de uma organização.

Um cliente com elevado grau de satisfação vale bastante mais para a instituição do que um cliente moderadamente satisfeito, uma vez que os clientes bastante satisfeitos são clientes menos propensos a mudar para a concorrência mesmo que esta, circunstancialmente, lhes apresente melhor oferta.

O trabalho que aqui se apresenta teve como principal objectivo estudar um tema cada vez mais emergente na actual literatura do marketing: a lealdade do consumidor face ao mercado Bancário.

Assim, na concretização deste trabalho foi desenvolvido um Modelo Preditivo de Data Mining. O Objectivo deste modelo consistiu na identificação dos clientes bancários com maior probabilidade para o abandono na instituição e sobretudo quais as razões subjacentes a este processo.

Este conhecimento dotará as instituições de competências que lhes permitirão definir estratégias direccionadas para a mitigação antecipada do risco de Abandono.

1.1) Relevância do Modelo de Abandono de Cliente na actualidade

O resultado de um estudo internacional efectuado pela consultora Everis (2010) e em 35 instituições bancárias de oito países, revela que 75% dos executivos do sector financeiro português e 60% a nível europeu reconhecem crescimento da predisposição dos clientes para mudarem de banco. Paradoxalmente, 60% dos investimentos estratégicos das entidades europeias estão focados na redução de custos no relacionamento com os clientes (Everis, 2010). O resultado deste estudo origina a necessidade de se analisar a problemática do Abandono, sobretudo as causas inerentes a este acto. A detecção destas causas constituirá aquisição de conhecimento precioso que permitirá à organização definir e implementar uma estratégia que tenda a mitigar essa causalidade, personalizar a relação com o cliente e reter os mais valiosos¹.

Para Lejeune (2001) a gestão do abandono consiste precisamente no desenvolvimento de técnicas que habilitem as empresas da capacidade de reter os seus clientes mais lucrativos apostando no incremento da sua lealdade. É importante manter os clientes satisfeitos mas é ainda mais importante fazê-lo de uma forma lucrativa². Se, subitamente, um cliente começar a agir num óbvio sentido de fricção com a instituição, a gestão deverá aplicar uma estratégia anti-abandono - que será necessariamente menos dispendiosa do que os custos de angariação de um novo cliente (Chu 2007). Será crucial proceder deste modo pois o custo estimado de atracção de um novo cliente é cinco vezes superior ao da respectiva manutenção (Kurtz; Clow, 1998). Por outro lado, estima-se que um incremento de 5% no esforço e concretização da retenção de clientes, pode redundar numa redução de cerca de 18% nos custos operacionais (Karakostas, 2005).

Inevitavelmente, o esforço de retenção terá de assumir particular importância na estratégia de todas as instituições que pretendam incrementar a sua eficiência, aumentar a sua rentabilidade e ultrapassar a concorrência. Construir relações de fidelidade entre a instituição e o cliente é uma forma de ganhar vantagem competitiva (McKenna, 1991; Reichheld, 1993).

1 Os clientes de maior valor serão aqueles que oferecem maior rentabilidade e potencialidade ao longo do

2 A conhecida regra 80/20, em que 20% dos principais clientes de uma instituição geram até 80% do respectivo lucro, aplicar-se-á a esta situação em grosso modo.

2. Enquadramento Teórico

A concorrência no mundo dos negócios reclama cada vez mais por dois factores críticos de sucesso e vantagem competitiva: velocidade³ e inovação⁴. Estes factores são pilares de uma equação aparentemente simples. Falhar a sua resolução, pode trazer dissabores difíceis de recuperar e ultrapassar. Grandes organizações foram já vítimas desse erro, a IBM foi disso exemplo, quando Thomas J. Watson, ex-presidente dessa companhia, profetizou nos anos 1940 que não haveria espaço para mais do que cinco computadores no mundo, em contraste com a visão de Bill Gates que ainda na década de 80 preconizou um computador pessoal em todas as casas, tendo essa ideia conduzido a Microsoft ao sucesso que hoje conhecemos.

Não obstante, velocidade e inovação são apenas efeitos de uma necessidade cada vez mais crescente nas organizações. Nesse contexto, Penna (2003) sugere que a informação e sobretudo o conhecimento (habilitado pelo bom uso da informação), tem papel preponderante na tomada de decisão táctica e estratégica das organizações.

Conclui-se portanto que, informação e conhecimento, são as matérias-primas que alimentam a equação da velocidade e inovação: informações sobre o mercado, sobre os concorrentes, sobre patentes, sobre a própria empresa. Todas estas informações produzem conhecimento sobre tendências, previsões mais assertivas, estudos de sazonalidade, participação de mercado, comparação de preços, entre outros.

Neste contexto, os sistemas de informação representam um papel deveras importante nas organizações.

2.1) Sistemas de Informação

Um sistema de informação pode ser definido como um conjunto organizado de elementos, pessoas, dados, actividades, recursos materiais em geral. Estes elementos interagem entre si para processar informação e divulgá-la de forma adequada em função dos objectivos de uma organização.

³ Na identificação de ameaças e oportunidades.

⁴ Não apenas na capacidade de corresponder às necessidades do consumidor mas principalmente na capacidade de o surpreender e de superar as suas expectativas.

Numa perspectiva empresarial, os sistemas de informação podem classificar-se de diversas formas. Existem, por exemplo, sistemas de informação transaccional⁵, sistemas de automatização⁶, sistemas de gestão operacional⁷, sistemas especializados⁸ e sistemas de informação estratégicos ou de suporte à decisão⁹.

As tecnologias de informação e comunicações ganham assim, pela velocidade e pela inovação, uma particular preponderância na organização, pelo facto de apoiarem a mensuração destes benefícios¹⁰ e de materializarem os diversos sistemas de informação num único, comumente designado por sistema de informação informático.

2.2) Sistemas de Suporte à Decisão

Dentro desse cenário cresce a necessidade por sistemas informatizados de apoio à decisão, para armazenar grandes quantidades de dados e transformá-los em informação executiva para a formulação ou adequação de estratégia¹¹.

Um sistema de informação deve ser estratégico e contribuir para que uma organização possa alcançar os seus objectivos (Rowley, 1995).

Os Sistemas de Suporte à Decisão são a classe dos Sistemas de Informação que presta apoio à decisão estratégica, orientada ao assunto, recorrendo à análise de um conjunto de dados históricos de múltiplas fontes.

Um Sistema de Suporte à Decisão é um Sistema de Informação interactivo, flexível e adaptável, especialmente desenvolvido para apoiar a solução de um problema de gestão não estruturado e aperfeiçoar a tomada de decisão. Este sistema utiliza dados, prevê um interface amigável e permite ao decisor ter a sua própria percepção (Turban, 1998). Estes sistemas necessitam de dados para processar informação. Neste contexto, torna-se pertinente efectuar uma rápida abordagem ao conceito de Data Warehouse.

5 Sistemas que gerem a informação referente às transacções que têm lugar numa empresa.

6 Aplicações que ajudam a melhorar a eficiência de tarefas repetitivas do trabalho administrativo.

7 Sistemas que apoiam a resolução de diversos problemas correntes do quotidiano.

8 Sistemas que emulam o comportamento de um especialista numa área concreta.

9 Estes sistemas sustentam a análise das diversas vertentes de negócio para apoio à tomada de decisões.

10 A utilidade e a necessidade desse sistema será percebida, se estes benefícios puderem ser quantificados e considerados relevantes (Fitzgerald, 1998).

11 Agrupando-os, resumindo-os, permitindo análises por diversas dimensões.

2.3) Data Warehouse

Um Data Warehouse pode ser definido como um conjunto de dados baseado em assuntos, integrado, não volátil e variável em relação ao tempo, de apoio às decisões de gestão (Inmon, 1995) ou, de uma forma mais simples mas não menos concisa, como uma cópia de dados transaccionais especificamente estruturada para análise e pesquisas (Kimball, 2002).

O Data Warehouse é pois um repositório devidamente organizado e vocacionado para apoiar os decisores. Este repositório contém todo o histórico da organização sobre um determinado assunto, proveniente das mais diversas fontes distribuídas da organização.

2.4) Gestão da Relação com os Clientes

Sendo parte do contexto deste trabalho o estudo comportamental do cliente (particularmente nas situações de abandono), torna-se imperativo abordar a relevância da componente de Gestão da Relação com o Cliente¹² numa organização.

A SAS conceptualiza o CRM como sendo um sistema de informação que aumenta a qualidade do relacionamento com o cliente, promovendo um incremento na retenção dos clientes em diversas vertentes:

- Suporta modelos preditivos que auxiliam os bancos a detectar situações de propensão ao abandono, as respectivas causas e como proceder nessas situações;
- Promove um novo nível de customização na oferta de serviços e nas iniciativas de marketing que fomentam a lealdade do cliente;
- Aporta elevado valor à interacção de e com o cliente¹³ pela consistência na informação partilhada por todos os canais de contacto com o cliente;

Para a Gartner, o CRM é uma estratégia de negócio com resultados que optimizam a rentabilidade, o lucro e a satisfação do cliente, organizando-se em torno de segmentos de clientes, promovendo comportamentos que satisfaçam os clientes e implementando processos de negócio centrados nos clientes centrados nos clientes. As tecnologias associadas ao CRM devem permitir um maior conhecimento dos clientes, maior acesso

¹² Também designado por CRM (Customer Relationship Manager)

¹³ Incrementando o nível de satisfação.

pelos clientes, interações mais eficientes e integração através de todos os canais de contacto com os clientes e funções de back-office das empresas (Ferrão, 2009).

Estas definições concentram efectivamente todos os pontos essenciais relacionados com o tópico, salientando a pertinência da interactividade bidireccional da comunicação de e para o cliente, preconizando uma dinâmica comportamental da organização centrada na personalização da relação com o cliente de modo a aumentar o seu grau de satisfação, abordando a importância do suporte tecnológico e da necessidade de integração da informação recolhida em todos os canais de contacto com o cliente¹⁴.

Independentemente da eventual arquitectura distribuída que sustente parte significativa dos sistemas de informação de uma organização é imperativo integrar e centralizar toda a comunicação com o cliente.¹⁵

14 Presenciais ou não presenciais.

15 O exemplo do cliente que, mudando de residência, comunica ao banco a respectiva alteração, verificando, mais tarde, que continua, assim mesmo, a receber correspondência na anterior morada, das seguradoras e de outras empresas do grupo com quem mantém relações - é infelizmente um clássico que mostra alguma ineficiência na utilização de um CRM por parte de grande parte das organizações.

2.5) Data Mining

A compreensão da informação recolhida sobre o cliente transforma-se em conhecimento com o apoio do Data Mining – conjunto de técnicas automáticas, usadas para explorar exaustivamente e descobrir relacionamentos complexos num grande conjunto de dados (Moxon, 1998).

O SAS define Data Mining como o processo de Amostragem, Exploração, Modificação e Modelagem de grandes volumes de dados para identificar padrões previamente desconhecidos, que possam ser utilizados como uma vantagem de negócio. O Data Mining é um processo que permite desenvolver e implementar dois tipos de modelos analíticos: (1) Modelos Comportamentais (ou Descritivos)¹⁶ e (2) Modelos Preditivos¹⁷.

2.5.1) Modelos Analíticos Preditivos

Os modelos Preditivos de Data Mining podem ser classificados nos seguintes tipos¹⁸:

- Retenção (ou Abandono)¹⁹;
- Cross-Sell²⁰;
- Up-Sell²¹;
- Next Best Offer²²;
- Análise de Sobrevivência/Duração²³;
- CLTV (Customer Lifetime Value)²⁴;

16 Os Modelos Comportamentais procuram identificar e analisar o comportamento dos clientes (ou do objecto de estudo em questão).

17 Os Modelos Preditivos visam detectar as propensões para determinados eventos e explicar o motivo que leva aos mesmos.

18 Dependendo da fase do ciclo de vida dos clientes.

19 Os modelos de Retenção de clientes permitem identificar propensões ao cancelamento ou abandono num determinado período, identificando as razões subjacentes ao mesmo.

20 Permite identificar os clientes que apresentam maior tendência à compra de determinados produtos ou serviços;

21 Modelos que se focam na probabilidade de aquisição de produtos de maior valor;

22 Procuram a definição da próxima melhor oferta a apresentar a cada um dos clientes;

23 Permite identificar e estimar qual o período de tempo médio de ocorrência de um determinado evento

Este trabalho final de curso assenta sobre o desenvolvido de um Modelo Preditivo de Retenção ou Abandono.

2.5.2) Técnicas utilizadas

Para o desenvolvimento dos modelos preditivos é possível utilizar diversas técnicas e parametrizações. As técnicas mais frequentes são:

- Regressões Lineares²⁵ (não utilizada neste trabalho);
- Regressões Logísticas: As regressões logísticas procuram descrever as probabilidades associadas aos valores da variável Target. Esta técnica é tipicamente utilizada em modelos preditivos que pretendam aferir probabilidades de concretização de um determinado evento categórico;
- Árvores de Decisão: As árvores de decisão baseiam-se na construção de um conjunto de regras proporcionais²⁶ que permitem explicar a variável Target, residente na raiz da árvore. Trata-se de uma técnica privilegiada para explicar as causas de um determinado evento, razão pela qual é tão utilizada;
- Redes Neurais²⁷ (não utilizada neste trabalho);
- Modelos de Mistura: técnica de refinação que produz um modelo de combinação de dois ou mais modelos cujo resultado é similar produzindo normalmente melhores taxas de ajustamento.

24 Preconizam a obtenção de conhecimento sobre o valor futuro dos clientes na organização, ao longo do seu ciclo de vida.

25 As regressões lineares visam a previsão do valor médio da variável dependente tendo em conta os valores das variáveis explicativas. Esta técnica é aplicada em modelos preditivos que visem quantificações, ou seja, prever e explicar eventos intervalares.

26 Regras do estilo “if... then... else”.

27 As redes neurais constituem uma estrutura onde existe uma camada de nós de entrada, outra de nós de saída, podendo ter ainda uma ou mais camadas de nós intermediários que realizam o processamento dos dados de entrada proporcionando um resultado na saída. Tipicamente este método produz resultados muito eficientes e ajustados quando se pretende exclusivamente prever um evento. Contudo, os processamentos gerados por esta técnica são considerados “opacos”, na medida em que não se pode interpretar claramente quais os factores que originaram determinado evento e por esta razão não aplicámos esta técnica no desenvolvimento deste trabalho

3. Método

3.1) Introdução

Este projecto teve como principal objectivo a construção de um modelo preditivo dotado da capacidade de detectar preventivamente situações de abandono de clientes na banca. Conceptualmente, no contexto deste trabalho, considera-se como abandono a situação de um cliente com status activo nos últimos 8 meses e que nos próximos 3 meses consecutivos altera esse mesmo status para inactivo.

Foram utilizados dados reais de uma instituição financeira portuguesa que por razões de confidencialidade não podem ser divulgados.

Tomando como referência o mês de Setembro de 2009, consideraram-se elegíveis os clientes com idade superior a 19 anos, que não sejam colaboradores da instituição, que estejam activos e não tenham falecido.

3.2) Metodologia

A abordagem seguida para a concretização do projecto foi composto por 7 etapas distintas e sequenciais, divididas por 3 fases (Definição, Planeamento e Execução):

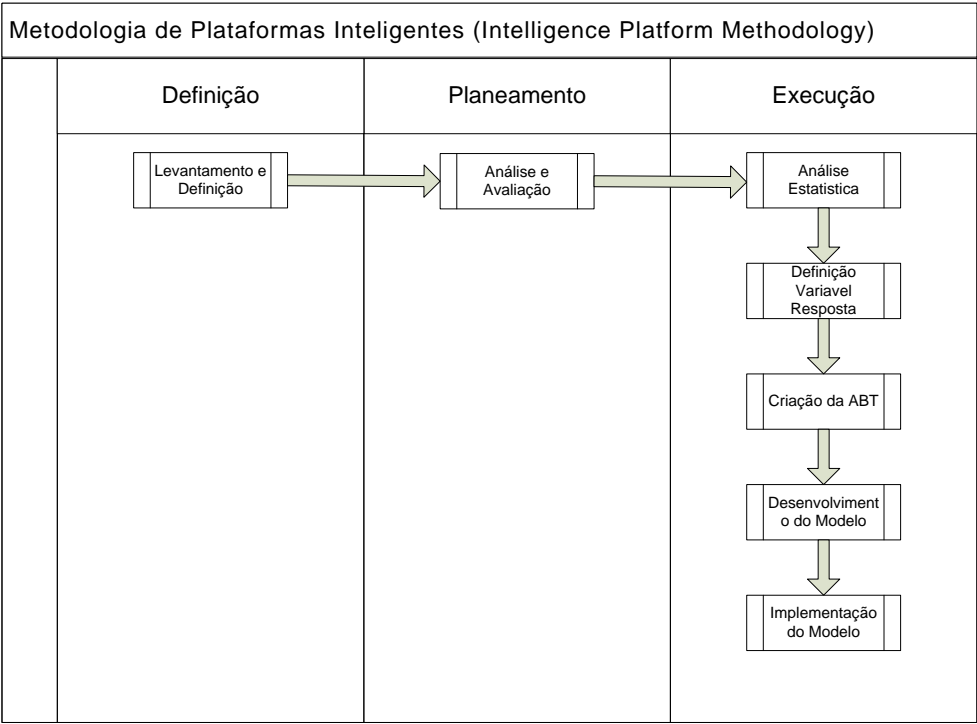


Figura 1: Metodologia de Plataformas Inteligentes

1. Na fase de Definição foi feita uma avaliação da situação actual do cliente e uma definição dos requisitos do modelo a desenvolver;
2. A segunda fase, Planeamento, consistiu numa análise introdutória à disponibilidade dos dados existentes e possíveis adequações²⁸. Esta fase contemplou também a decisão da escolha do software a usar no projecto:
 - SAS Enterprise Miner (desenvolvimento do Modelo Preditivo)
 - SAS Enterprise Guide e MSFT Excel (elaboração de gráficos e estatísticas)
3. Já na fase de Execução, foi efectuada uma Análise Estatística prévia, de forma a analisar se os dados fornecidos pela entidade financeira eram de qualidade e possíveis de se utilizar²⁹;
4. Na etapa de Definição da Variável Resposta foi estabelecido o critério de abandono do cliente do banco: conceptualmente, num horizonte temporal de 11 meses, considerar-se-á situação de abandono aquelas em que cliente se mantenha com status activo nos últimos 8 meses e transite para inactivo nos próximos 3 meses.

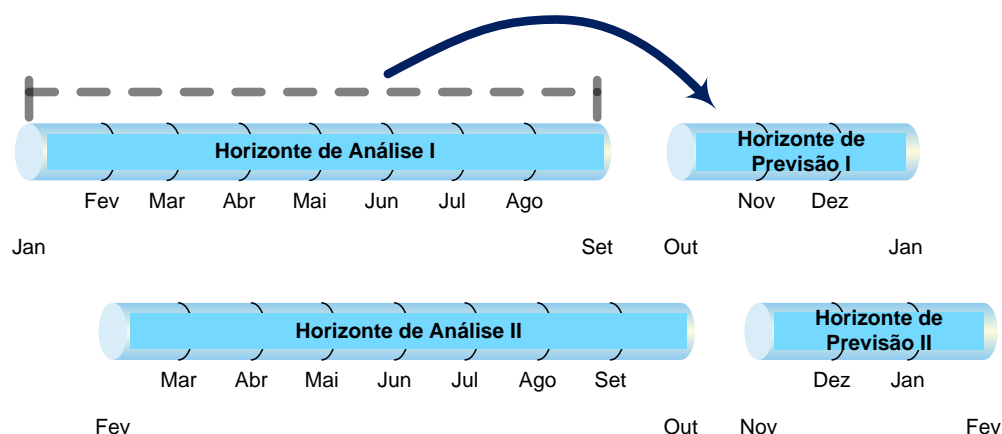


Figura 2 – Horizonte temporal para abandono

5. Na etapa de Criação da Analytical Base Table³⁰ foi efectuada uma selecção das variáveis e casos (clientes) elegíveis para o modelo em questão³¹.
6. A etapa de Desenvolvimento do Modelo consistiu no processo de desenho dos modelos de previsão, utilizando o processo SEMMA³², com posterior eleição do modelo final, tendo em conta as estatísticas de ajustamento do modelo.

28 A amostra piloto que serviu de base a esta análise continha um ano de dados de histórico para cada cliente.

29 Por exemplo, variáveis com uma grande quantidade de valores omissos não foram consideradas.

30 ABT consiste numa tabela que contem os registos dos clientes.

31 Tabela esta que foi entregue posteriormente pela própria entidade

32 SEMMA é acrónimo de Sample, Explore, Model, Modify, Asset pela SAS Institute Inc.

7. Na etapa de Implementação do Modelo, é efectuada a passagem para Produção do modelo eleito. A qualidade do Modelo é validada periodicamente ao longo do tempo e caso haja um fraco ajustamento do modelo o mesmo deverá ser revisto.

3.3) Descrição dos Dados

A ABT utilizada é estruturada por 290 variáveis (ver Anexo 1³³) e 972706 registos de clientes, dos quais 8027 (0,83%) representam abandonos e os restantes 964679 (99,17%) situações de não abandono.

A variável resposta para o modelo de abandono de clientes produz efeitos num período temporal de 3 meses, segundo os critérios de negócios definidos para o abandono.

A Target Abandono, sendo binária, poderá assumir um de dois valores:

1	Se o cliente teve status activo nos últimos 8 meses e nos próximos 3 meses consecutivos esteve com status inactivo (abandono).
0	Se o cliente teve status activo nos últimos 8 meses e nos próximos 3 meses consecutivos continuou com status activo (não abandono).

Tabela 1– Significado dos valores da variável target

A percentagem de casos de abandono é reduzida, razão pela qual torna-se necessário efectuar uma abordagem de *Oversampling*³⁴. O objectivo é captar os comportamentos dos indivíduos que Abandonam o Banco.

Os escalões etários mais representativos compreendem o intervalo entre 40 e 54 anos de idade (36,69%) e entre os 30 e 39 anos (26,53%); os escalões etários de menor representatividade correspondem aos mais jovens com idades entre 18 e 24 anos (3,32%) e entre 25 e 29 anos (6,66%).

No que diz respeito ao género, 37,94% dos clientes são do sexo feminino e os remanescentes 62,06% do sexo masculino.

Os clientes são maioritariamente casados ou em união de facto (65,56%) embora os solteiros também abarquem uma percentagem significativa (22,77%) e praticamente todos de nacionalidade portuguesa (97,78%).

Relativamente ao potencial bancário dos clientes conclui-se que: (1) é maioritariamente baixo (55,81%); (2) 9,15% tem potencial bancário Elevado (9,15%) e (3) 0,80% Muito Elevado.

³³ Refere a descrição de variáveis da ABT.

³⁴ Redimensionar a diferença entre a percentagem de evento e de não evento – explicado na secção 3.4.1.

Já o potencial Sócio-Económico dos clientes é: (1) maioritariamente Médio (44,36%); (2) 14,38% tem potencial Elevado (14,38%) e (3) 3,64% Muito Elevado.

O cruzamento destas duas últimas dimensões leva a concluir que o potencial Sócio-Económico dos clientes mais valiosos ($9,15\% + 0,80\% = 9,95\%$) não tem correspondência directa no Potencial Bancário ($14,38\% + 3,64\% = 18,02\%$) o que indicia que a lealdade à instituição não é plena, facto que, por sua vez, constitui uma boa oportunidade de exploração.

Numa perspectiva geográfica, os clientes são maioritariamente residentes em concelhos Metropolitanos (43,38%) ou Urbanos (36,62%) em concelhos considerados como sendo de Riqueza Média (45,28%), Média Alta (28,54%) ou mesmo Ricos (19,25%).

Para maior detalhe na distribuição Sócio-Demográfica, consultar tabelas do Anexo 2³⁵.

³⁵ Anexo que traduz a distribuição de valores das variáveis sócio-demográficas.

3.4) Desenvolvimento do Modelo Preditivo

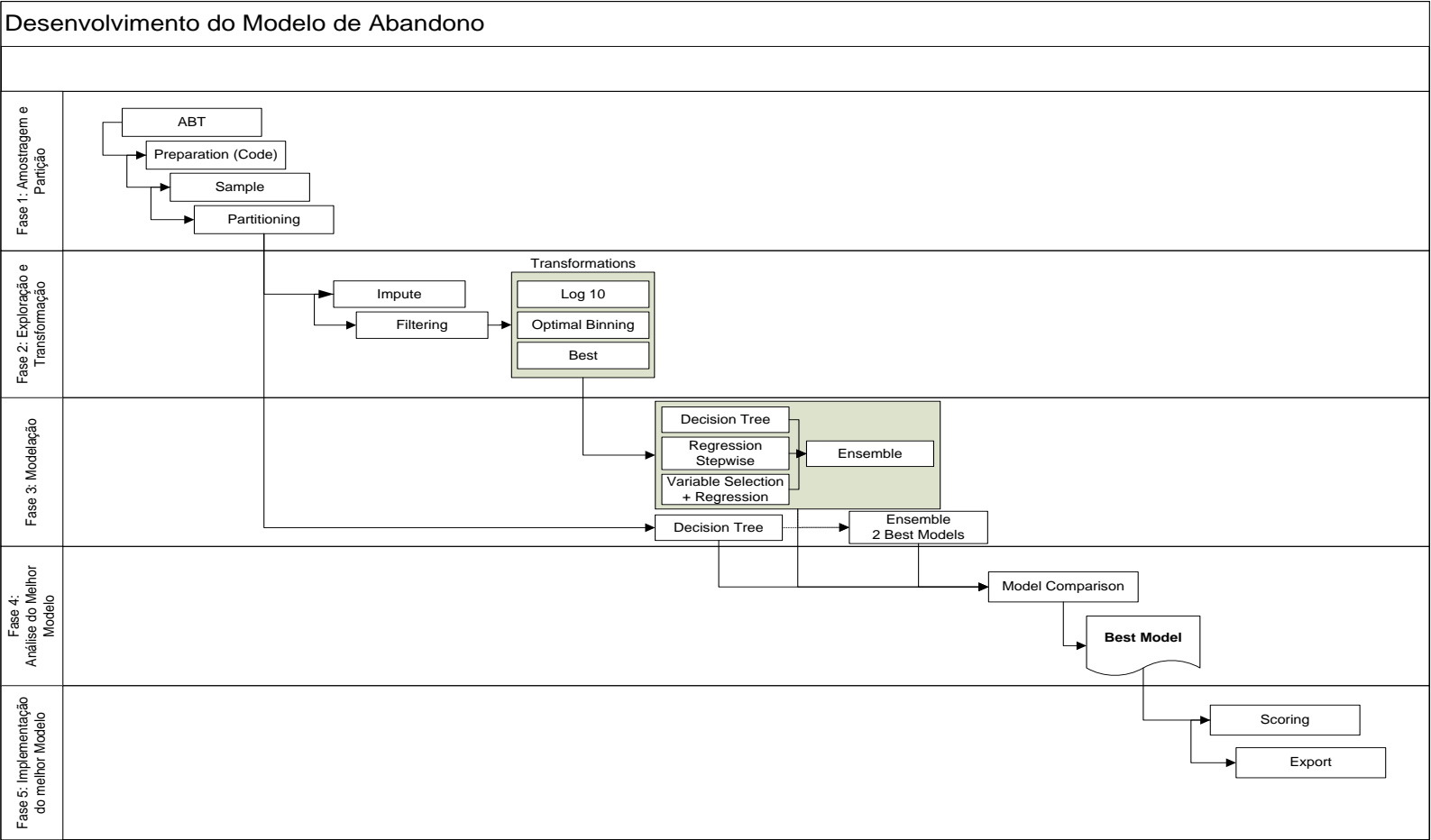


Figura 3: Diagrama de Desenvolvimento do Modelo Preditivo.

O processo de desenvolvimento de um modelo Data Mining é um processo de complexidade elevada e que requer atributos de curiosidade, perseverança e resiliência da equipa envolvida.

No decurso deste projecto foram testadas inúmeras hipóteses de modelação, umas rapidamente rejeitadas, outras posteriormente sujeitas a um processo de afinação e melhoria contínua, naturalmente em busca dos melhores resultados possíveis.

O Modelo Preditivo de Abandono de Clientes foi desenvolvido tendo em conta 5 grandes fases. Cada fase tem objectivos distintos:

1. Amostragem e partição
2. Exploração e transformação
3. Modelação
4. Comparação e Análise melhor modelo
5. Implementação do melhor modelo

3.4.1.) Amostragem e partição

Inicialmente importou-se a ABT e procedeu-se à selecção do papel e do tipo que cada variável assume no Modelo. De seguida, após a importação da ABT, foi realizado um pré-tratamento de algumas situações de valores nulos que preferimos classificar directamente com códigos específicos para situações indeterminadas. Esta operação foi efectuada por aplicação directa de **SAS Code**³⁶.

A ABT inicial era constituída por 99,17% de registos referentes a situações de não abandono. Com o objectivo de facilitar a análise, este forte desequilíbrio entre a percentagem de casos de Abandono e de casos de Não Abandono foi atenuado no modelo por um balanceamento por oversampling aplicado num nó de **Sample**³⁷. A aplicação desta técnica provocou uma alteração no universo dos dados. Assim, a amostra passou a ser constituída por igual número de registos - 8027 - de abandonos e não abandonos, num total de 16054.

Com o intuito de garantirmos uma boa capacidade de generalização do modelo a desenvolver a novos dados efectuou-se uma separação dos dados, através do nó de **Data Partition** do SAS Enterprise Miner, em dois conjuntos de dados: (1) amostra/conjunto

³⁶ Ver Anexo 3: SAS Code utilizado.

³⁷ Ver Anexo 4: Node Sample.

de treino e (2) amostra/conjunto de validação. Assim, partindo da amostra de 16054 registos, foi utilizado este nó com o intuito de separar 70% dos dados (5618 registos) para a amostra de treino e os restantes 30% (2409 registos) para a amostra de validação³⁸.

3.4.2.) Exploração e transformação

A partir do nó Data Partition foram efectuadas explorações para aferir os relacionamentos da variável Target com as variáveis sócio-demográficas³⁹. Verificou-se maior tendência para o abandono nos seguintes grupos de clientes: idades jovens (entre 18 e 24 anos), potencial bancário e sócio-económico baixo ou muito baixo, com habilitações literárias ao nível do Mestrado. Ao invés, os grupos de potencial bancário Médio e potencial sócio-económico Médio ou Elevado apresentaram maior tendência para a permanência do que para o abandono.

A análise aos resultados a um nó **Stat Explore** colocado a seguir ao nó de particionamento revelou uma existência considerável de valores omissos que foram posteriormente tratados por intermédio de um nó **Impute** colocado na sequência do nó de particionamento⁴⁰. O tratamento foi efectuado por intermédio de uma Tree⁴¹ tanto para as variáveis intervalares como para as de categóricas. O valor aplicado por defeito nas variáveis numéricas foi “0”; para as variáveis de texto foi aplicado o valor “D”⁴².

A fim de homogeneizar os dados, atenuando eventuais situações de *outliers*, foi aplicado um nó **Filter**⁴³ posterior ao nó de Impute, aplicando o método de exclusão de Percentis Extremos nas variáveis intervalares⁴⁴ e o default Rare Values para as categóricas⁴⁵.

Posteriormente, para melhor ajustamento dos dados e devido à assimetria da distribuição de algumas variáveis, foram aplicados, através do nó de **Transform**

³⁸ Ver Anexo 5: Node Data Partition.

³⁹ Ver gráficos no Anexo 7: Relacionamento da variável Target com as Sócio-Demográficas.

⁴⁰ Em paralelo com o nó Stat Explore.

⁴¹ Método particularmente vocacionado para tratamento de valores omissos.

⁴² Ver Anexo 6: Node Impute.

⁴³ Ver Anexo 8: Node Filter

⁴⁴ Os valores das variáveis intervalares que fossem acima do valor do percentil extremo eram filtrados.

⁴⁵ Sempre que existirem valores raros numa classe da variável esta classe é agrupada com uma outra.

Variables⁴⁶ três tipos de transformações distintas⁴⁷: (1) logaritmo de cada variável (Log 10); (2) transformação do tipo Binning⁴⁸ e (3) transformação do tipo Best Power (Best)⁴⁹.

As transformações foram testadas e afinadas ao longo do processo de desenvolvimento do modelo, sendo estas que, em cada tipo, permitiram a obtenção de melhores resultados sobre o conjunto de testes e também sobre o conjunto de validação.

3.4.3.) Modelação

Antes de iniciar o processo de modelação é necessário aplicar um procedimento de selecção de variáveis com as quais o modelo irá trabalhar. Isto acontece já que a quantidade de variáveis de input a utilizar é elevada. Existem diversas técnicas possíveis de se utilizar para reduzir a dimensionalidade dos dados:

(1) Nó de **Variable Selection**⁵⁰

(2) Nó de **Decision Tree**⁵¹

(3) Nó de **Regressão**⁵²

Após a selecção automática de quais as variáveis explicativas que, à partida, são mais relevantes para prever e compreender a variável Target é possível modelar o evento em estudo – Abandono de Clientes no Banco. Para a modelação foram utilizados diversos métodos distintos, aplicados após as transformações e/ou selecções de variáveis (nó

⁴⁶ Ver Anexo 9: Transformation Node.

⁴⁷ Sendo que cada método tem uma própria transformação específica.

⁴⁸ Este método permite criar n classes das variáveis, de forma a optimizar a relação com a Target.

⁴⁹ Onde a transformação é escolhida de forma a minimizar a variância entre cada variável e a Target.

⁵⁰ Este nó ajuda a reduzir o número de variáveis de input, seleccionando e rejeitando aquelas que não estão relacionadas, ou seja, que pouca significância estatística tem com a Target;

⁵¹ O algoritmo de definição das árvores de decisão calcula a medida de importância de cada variável e aplica-lhes um valor entre 0 e 1, sendo 1 o valor de maior importância. Este método tem a vantagem de processar rapidamente quantidades elevadas de variáveis e de poder efectuar autonomamente tratamento de omissos;

⁵² Que pode incluir, ele próprio, um processo de selecção de variáveis: Stepwise. Este é um método de selecção progressiva ao qual se junta, após a inclusão de uma nova variável, um passo em que se testa a significância de todas as variáveis incluídas no modelo e se retiram aquelas que não forem significativas. Este método de selecção é lento quando processa quantidades elevadas de variáveis, embora muito eficiente.

especifico de **Variable Selection** com selecção prévia de variáveis por *R and Chi-Square*⁵³.

Os métodos utilizados foram: (1) Árvore de Decisão; (2) Regressão Logística utilizando como método de selecção de variáveis o método de Stepwise e (3) Regressão Logística com todas as variáveis seleccionadas anteriormente através do nó de Variable Selection. A cada um destes conjuntos foi por sua vez aplicado um Modelo de Mistura. A tabela 2 permite identificar todas as transformações analisadas e técnicas preditivas utilizadas nos diversos testes desenvolvidos.

Transformação	Modelo	Modelo de Mistura
Log 10	Árvore de Decisão Log 10	Mistura Log10
	Regressão Stepwise Log10	
	Variable Selection + Regressão Log10	
Optimal Binning	Árvore de Decisão Optimal Binning	Mistura Optimal Binning
	Regressão Stepwise Optimal Binning	
	Variable Selection + Regressão Optimal Binning	
Best	Árvore de Decisão Best	Mistura Best
	Regressão Stepwise Best	
	Variable Selection + Regressão Best	

Tabela 2– Correspondência de Modelos e Transformações

Para além destes modelos, e aproveitando a sua capacidade em lidar bem com valores omissos, foi aplicado um nó de **Decision Tree**, logo após o nó de Data Partition.

Finalmente, a terminar o processo, adicionámos um modelo de mistura aplicado aos dois melhores modelos até então detectados. Isto porque ambos os modelos não apresentavam grandes diferenças em termos de ajustamento.

O projecto incluiu portanto as seguintes técnicas preditivas (tabela 3).

⁵³ Ver Anexo 10: Node Variable Selection.

##	Modelo	Designação Final
1	Árvore de Decisão Log 10	Dec Tree Log10
2	Regressão Stepwise Log10	Reg Stepwise Log10
3	Variable Selection + Regressão Log10	Reg VS Log10
4	Mistura Log10	Ensemble Log10
5	Árvore de Decisão Optimal Binning	Dec Tree OptBin
6	Regressão Stepwise Optimal Binning	Reg Stepwise OptBin
7	Variable Selection + Regressão Optimal Binning	Reg VS OptBin
8	Mistura Optimal Binning	Ensemble OptBin
9	Árvore de Decisão Best	Dec Tree Best
10	Regressão Stepwise Best	Reg Stepwise Best
11	Variable Selection + Regressão Best	Reg VS Best
12	Mistura Best	Ensemble Best
13	Árvore de Decisão sem Impute	Dec Tree not Imputed
14	Mistura 2 melhores modelos	Ensemble 2 Best Models

Tabela 3 - Designação final dos modelos

3.4.4.) Análise do melhor modelo

A análise do melhor modelo foi efectuada com base ao recurso de um nó de **Model Comparison** que⁵⁴ avalia diversas medidas de ajustamento de cada um dos modelos e com base nessa avaliação cria um ranking de modelos ordenado pelos melhores resultados. Assim, este nó existente no SAS Enterprise Miner permite analisar e comparar todas as técnicas utilizadas e seleccionar o melhor modelo. A selecção do modelo (Y) baseia-se em duas medidas de ajustamento do modelo:

- ‘Misclassification Rate’⁵⁵;
- ‘Roc Index’⁵⁶;

A tabela 4 ilustra os resultados destas estatísticas tanto para o conjunto de treino como para o conjunto de validação.

⁵⁴ Independentemente da dimensão da amostra ou do tipo de modelo testado.

⁵⁵ Esta medida é adequada para targets binárias e exprime a taxa de má classificação, ou seja, classificar o evento (1: abandono) como não evento (0: não abandono) ou ainda, classificar o não evento (0) como evento (1). Quanto menor esta medida melhor a performance do modelo;

⁵⁶ Medida muito aplicada na avaliação da performance de modelos de resposta binária e está associado ao poder discriminante de um teste de diagnóstico. Assim, exprime o grau de precisão do evento e do não evento. Maiores valores deste índice representam maior poder discriminante, i.e., melhor performance do modelo;

Selected Model	Model Description	Valid: Misclassification Rate	Train: Misclassification Rate	Valid: Roc Index	Train: Roc Index
Y	Ensemble 2 Best Models	0.1461187214611872	0.14982361485785434	0.932	0.928
	Reg Stepwise Best	0.14757160647571607	0.14992737082382235	0.931	0.929
	Reg Stepwise OptBin	0.1477791614777916	0.1517949782112471	0.931	0.926
	Ensemble Best	0.1552511415525114	0.1553226810541606	0.929	0.923
	Dec Tree not Imputed	0.15545869655458697	0.15236739053043788	0.901	0.899
	Ensemble OptBin	0.15566625155666253	0.1550114131562565	0.927	0.922
	Ensemble Log10	0.15960979659609797	0.17088607594936708	0.92	0.908
	Dec Tree Best	0.15960979659609797	0.15999169952272255	0.907	0.895
	Reg VS Log10	0.16251556662515568	0.1752438265200249	0.918	0.906
	Reg Stepwise Log10	0.16293067662930677	0.17918655322681054	0.917	0.903
	Reg VS Best	0.16542133665421338	0.16393442622950816	0.921	0.915
	Reg VS OptBin	0.1656288916562889	0.16798090890226192	0.919	0.914
	Dec Tree Log10	0.16583644665836447	0.16642456941274122	0.899	0.889
	Dec Tree OptBin	0.16625155666251556	0.16715086117451752	0.889	0.878

Tabela 4 - Avaliação dos Modelos

De acordo com estas duas métricas, o melhor resultado é obtido pelo modelo “Ensemble 2 Best Models” que resultou da mistura entre os modelos “Reg Stepwise Best” e “Reg Stepwise OptBin”⁵⁷. Isto porque é o modelo com menor valor de ‘Misclassification Rate’ e maior valor de ‘Roc Index’ no conjunto de validação⁵⁸.

Analisando as diferenças entre o conjunto de treino e o conjunto de validação verifica-se a existência de pequenas variações, o que confere, à partida, maior consistência da análise e significa que o modelo terá uma boa capacidade de ajustamento a novos dados.

Os indicadores registados na *Event Classification Table* (em %) destacam o modelo “Ensemble 2 Best Models” como o melhor modelo, apontando em seguida para os modelos “Reg Stepwise Best” e “Reg Stepwise OptBin”. Nestes modelos, a detecção de situações verdadeiras⁵⁹ é claramente superior aos valores obtidos nos restantes modelos.

⁵⁷ E que ocupam a 2ª e 3ª posição do ranking de ordenação.

⁵⁸ Visto ser o conjunto de dados sobre os quais o modelo não foi desenvolvido.

⁵⁹ Isto é: situações em que se prevê o evento de abandono e em que ele efectivamente se concretiza – True Positive – e as situações em que se prevê o evento de não abandono e em que ele efectivamente não se concretiza – True Negative.

Model Description	Data Role	False Negative %	True Negative %	False Positive %	True Positive %	Sum True %
Ensemble 2 Best Models	VALIDATE	5,1%	40,5%	9,5%	44,9%	85,4%
Reg Stepwise Best	VALIDATE	6,1%	41,3%	8,7%	43,9%	85,2%
Reg Stepwise OptBin	VALIDATE	6,1%	41,3%	8,7%	43,9%	85,2%
Ensemble 2 Best Models	TRAIN	4,9%	36,9%	10,1%	48,1%	85,0%
Reg Stepwise Best	VALIDATE	5,6%	37,6%	9,4%	47,4%	85,0%
Reg Stepwise OptBin	TRAIN	5,7%	37,4%	9,5%	47,4%	84,8%
Dec Tree not Imputed	VALIDATE	3,9%	38,7%	11,3%	46,1%	84,8%
Ensemble Best	VALIDATE	6,6%	41,1%	8,9%	43,4%	84,5%
Ensemble Best	TRAIN	6,0%	37,4%	9,6%	47,1%	84,5%
Dec Tree not Imputed	VALIDATE	4,5%	39,0%	11,0%	45,5%	84,5%
Ensemble OptBin	VALIDATE	6,7%	41,1%	8,9%	43,3%	84,4%
Dec Tree Best	TRAIN	6,9%	41,0%	9,0%	43,1%	84,0%
Ensemble Log10	TRAIN	6,3%	40,3%	9,7%	43,7%	84,0%
Dec Tree Best	VALIDATE	6,8%	37,7%	9,2%	46,3%	84,0%
Reg VS Log10	TRAIN	6,3%	40,0%	10,0%	43,7%	83,7%
Reg Stepwise Log10	TRAIN	7,1%	40,8%	9,2%	42,9%	83,7%
Reg VS Best	TRAIN	6,2%	36,8%	10,2%	46,8%	83,6%
Reg VS Best	TRAIN	7,2%	40,7%	9,3%	42,8%	83,5%
Reg VS OptBin	TRAIN	7,4%	40,8%	9,2%	42,6%	83,4%
Dec Tree Log10	VALIDATE	6,3%	39,7%	10,3%	43,7%	83,4%
Dec Tree OptBin	VALIDATE	7,1%	40,5%	9,5%	42,9%	83,4%
Dec Tree Log10	TRAIN	6,0%	36,3%	10,7%	47,1%	83,4%
Dec Tree OptBin	TRAIN	7,0%	37,2%	9,7%	46,0%	83,3%
Reg VS OptBin	VALIDATE	6,7%	36,8%	10,1%	46,4%	83,2%
Ensemble Log10	VALIDATE	6,0%	35,8%	11,1%	47,1%	82,9%
Reg VS Log10	TRAIN	6,1%	35,6%	11,4%	46,9%	82,5%
Reg Stepwise Log10	VALIDATE	6,7%	35,8%	11,2%	46,3%	82,1%

Tabela 5 - Event Classification Table (Validation)

3.4.5.) Implementação do melhor modelo

Após a eleição do melhor modelo, importa disponibilizar o projecto em produção, para que periodicamente o mesmo possa ser processado sobre novos clientes (Scoring).

Para tal inseriu-se, após o modelo vencedor, o Nó de **Score**. Este nó permite disponibilizar todo o código subjacente à criação do modelo e aplicar o mesmo a novos clientes.

Após o nó de Scoring, foi ainda colocado um nó de **SAS Code**⁶⁰ que permitirá exportar para uma tabela específica fisicamente, os dados que identificam as situações propensas a abandono. Esta tabela permitirá à entidade bancária actuar directamente sobre estas potenciais ocorrências de abandono, de modo a precaver as situações mais críticas.

⁶⁰ Ver: Anexo 3: SAS Code utilizado.

4. Resultados

Após o desenvolvimento e implementação do modelo de Abandono de Clientes obtém-se, periodicamente como output, uma tabela que contém as probabilidades de abandono do Banco para cada cliente. Podemos considerar que clientes cuja probabilidade é superior a 50% são aqueles que efectivamente tendem a abandonar a Organização.

4.1) Identificação das variáveis mais relevantes para explicar o Abandono de Clientes no Banco.

Apesar de não ser difícil intuir que o “número de movimentos que um cliente realizou no mês actual” tem forte relevância no relacionamento com as situações de abandono⁶¹, tal como as Domiciliações de Ordenado e outras situações, este trabalho permitiu comprovar e quantificar cientificamente a relação entre as variáveis independentes e a variável dependente (Target), bem como a respectiva intensidade dessa relação.

O quadro ilustrado no Anexo 11⁶² contém o ranking de importância das variáveis de input no modelo vencedor. O parâmetro Chi-Square reflecte a relevância e associação da relação entre cada variável independente/explicativa e a variável Target – quanto maior o seu valor, mais significativa será essa relação.

4.2) Assessment Score Ranking

Da análise das tabelas ‘Assessment Score Ranking’⁶³ relativas a cada um dos modelos conclui-se que o modelo “Ensemble 2 Best Models” é o que apresenta melhores resultados, verificando-se que 98,75% das observações do percentil 5 correspondem a casos de evento, os quais apresentam, em média, valores de probabilidade superiores a 0,98913. Alargando o número de observações até ao percentil 25 são ainda registados 95,6846% de casos de evento, os quais apresentam, em média, valores de probabilidade

⁶¹ Um valor reduzido ou nulo de movimentos implicará uma elevada probabilidade de abandono; na inversa, um número elevado de movimentos, implicará uma reduzida probabilidade de abandono.

⁶² Ranking de importância das variáveis.

⁶³ Anexo 12: Assessment Score Ranking.

superiores a 0,91107. Da observação da estatística lift podemos ainda verificar que no percentil 5 existe pelo menos 1,97 vezes mais casos de evento⁶⁴.

A análise do conteúdo da tabela *Assessment Score Ranking* tem uma vertente prática muito vincada. Se, por hipótese, um departamento de Marketing pretender encetar uma campanha que vise precaver situações de iminente abandono, poderá direcciona-la, por exemplo, apenas para o percentil 5, conseguindo desse modo abranger 98,75% de casos de clientes nessa situação. Em alternativa, poderá optar por direccionar a mesma campanha para todos os clientes até ao percentil 25, atingindo assim 95,68% de situações de iminente abandono.

Será fácil concluir que, esta informação será preciosa para a elaboração de uma matriz de Custo/Benefício que permita avaliar qual o custo óptimo que potencia uma maximização de benefícios.

Admitamos, a título de exemplo, que o valor de custo da campanha por cliente é de 5€ e que o benefício estimado da recuperação de cada cliente é de 25€; admitamos ainda que a expectativa de sucesso da campanha de marketing aponta para os 30% e que o universo da amostra é de 4818 clientes⁶⁵.

Assim, extrapolando estes dados para uma folha de cálculo⁶⁶, facilmente concluiríamos que o valor de retorno (benefício) da campanha é inferior aos respectivos custos a partir do percentil 75 e que o valor óptimo de maximização de benefícios é atingido se a campanha for direccionada para o percentil 45.

⁶⁴ Ver gráficos no Anexo 13

⁶⁵ Universo de Validação.

⁶⁶ Anexo 14: Extrapolação de Custo/Benefício a partir de tabela de *Assessment Score Ranking*

4.3) Acções de Marketing

As oportunidades que o resultado deste projecto suscita prendem-se naturalmente com o objectivo de retenção dos clientes cujo valor e potencial sócio-económico é mais elevado e que foram identificados como estando em situação de iminente abandono. Assim, são sugeridas as seguintes actuações, tendo em conta a respectivo segmento e grupo:

Segmento	Grupo	Acção
Potencial Sócio Económico elevado	Geral	Contacto personalizado do Gestor de Conta.
	Jovens	Vouchers Experiência (SPA, Gourmet, Aventura)
	Adultos	Duplicar plafond Cartão Crédito sem aumento de anuidade
	Seniores	Promoção para contas depósito a prazo, disponibilizando taxa especial acima do mercado para depósitos a prazo e contas poupança reforma.
Potencial Sócio Económico médio	Geral	Acção de marketing directo por email.
		Comunicação (email ou carta personalizada), dando a conhecer os destaques e vantagens de utilização de canais complementares (Web e Mobile).

Tabela 6 - Actuações sugeridas sobre os clientes

5. Conclusões e Trabalho Futuro

5.1) Conclusões

Este projecto visou o tema da problemática do abandono de clientes da banca, realçando a importância da manutenção de relações de fidelidade, sobretudo com os clientes mais valiosos.

Factores como Velocidade e Inovação são factores críticos de sucesso de uma organização, sendo a tecnologia, a informação e o conhecimento os pilares que sustentam esses dois vectores. Nesse âmbito, tivemos oportunidade de utilizar o Data Mining como processo de transformação de informação em conhecimento e de sublinhar a relevância do sistema de Gestão de Relacionamento de Clientes nas organizações como fonte de informação privilegiada do comportamento dos clientes e como catalisador de processos que visem a satisfação e participação interactiva desses mesmos clientes, incrementando assim a sua lealdade.

A componente técnica do projecto adoptou a metodologia que o SAS Institute aplica no desenvolvimento e implementação de projectos de Data Mining – Metodologia “Implementação de Plataformas Inteligentes” (IPI). Descrevemos integralmente o processo de modelação efectuado, bem como as diversas técnicas preditivas aplicadas.

Demonstrou-se posteriormente a fase de eleição do melhor modelo, bem como as razões que sustentaram essa decisão. Foi escolhido um Modelo de Mistura que combinou os dois melhores modelos até então obtidos, ambos de Regressão Logística (Stepwise), provenientes de dois processos de transformação de variáveis distintos.

Comentámos os resultados obtidos, identificando e quantificando o peso das variáveis mais relevantes na relação com a variável Target. As variáveis dependentes mais significativas foram: (1) os movimentos que o cliente realizou no mês actual⁶⁷; (2) o facto do Cliente ter Domiciliações de Ordenados depositados à ordem no mês actual⁶⁸; (3) quociente entre o volume Actual em recursos e o que ele teve nos últimos 9 meses⁶⁹;

⁶⁷ Um valor nulo ou reduzido estará associado a uma iminente situação de abandono.

⁶⁸ A não domiciliação de ordenado denuncia uma acentuada propensão ao abandono.

⁶⁹ Um valor baixo indica uma predisposição para abandono.

- (4) tempo que falta para o término do Crédito à Habitação da última conta a vencer⁷⁰ e
(5) saldo da conta de depósitos a prazo que o cliente tem no mês actual⁷¹.

Posteriormente, sugerimos eventuais acções de marketing e destacámos a importância da tabela Assessment Score Ranking, na vertente prática que poderá assumir, através de uma matriz de custo/benefício que dará indicações pertinentes sobre o alvo para o qual deverá ser direccionada a campanha de retenção.

A identificação das situações de abandono bem como das respectivas causas representam uma aquisição de conhecimento precioso para a organização que lhe permite proporcionar uma vantagem competitiva, muitas vezes decisiva, sobre os outros concorrentes.

5.2) Trabalho Futuro

Sugere-se que num projecto futuro sejam acolhidas as seguintes medidas:

- Integração de uma variável de input que identifique o segmento de cada cliente em conformidade com o seu valor para o banco, baseado no envolvimento com a entidade bem como com o seu potencial bancário e sócio-económico.
- Introdução de uma variável explicativa que revele o potencial futuro do cliente (CLTV⁷²).
- Inclusão de uma variável resposta para registar impacto das campanhas (respondeu/não respondeu) em cada cliente; esta possibilidade permitirá diferenciar futuras campanhas.

⁷⁰ Um valor reduzido indicará uma probabilidade de abandono.

⁷¹ Um montante escasso aponta para uma eventual situação de abandono.

⁷² Customer Lifetime Value.

6. Bibliografia

- ROSSI, Carlos A V.; SLONGO, Luiz A. Estado-da-arte e Proposição de um Método Brasileiro de Satisfação de Clientes. Anais do 20º Enanpad, 1997.
- EVERIS. Excelência na gestão de clientes – Estudo global de executivos de entidades financeiras, edição 2010
- LEJEUNE, M.A.P.M. Measuring the impact of data mining on churn management. *Electronic Networking Applications and Policy*, Vol. 11, No. 5, 2001
- CHU, B-H., Tsai, M-S. and HO, C-S. Toward a hybrid data mining model for customer retention', *Knowledge-Based Systems*, Vol. 20, 2007.
- KURTZ, David L., CLOW, Kenneth. E. *Services Marketing*. New York: John Wiley & Sons, 1998.
- KARAKOSTAS, B., Kardaras, D. and PAPATHANASSIOU, E. 'The state of CRM adoption by the financial services in the UK: an empirical investigation', *Information & Management*, Vol. 42, 2005.
- McKENNA, Regis, *Relationship Marketing: Successful Strategies for the Age of the Customer*. MA, Addison –Westly, 1991.
- PENNA, Rogerio Adriano Castelpogii, REIS JUNIOR, Alderico Sales dos. O Data Warehouse como Suporte à Inteligência de Negócio. VI SIMPOI - Simpósio de Administração da Produção, Logística e Operações Internacionais. São Paulo, 2003.
- REICHHELD, F.F., *Loyalty Based Management*, *Harvard Business Review*, 1993.
- SEWEEL, Carl, BROWN, Paul. *Customers for Life: How to Turn That One-Time Buyer Into a Lifetime Customer*, Crown Business, 2002.
- FITZGERALD, Guy. Evaluating information systems projects: a multidimensional approach. *Journal of Information Technology*, v. 13, n. 1, 1998.
- ROWLEY, Jennifer. Strategic information systems planning. *Information Services & Use*, v. 15, n. 1, 1995.
- TURBAN, Efraim e Aronson, JAY E., *Decision Support Systems and Intelligent Systems*, Prentice Hall, 1998.
- INMON, W. H. What is a Data Warehouse?, *Prism*, Volume 1, Number 1, 1995.
- KIMBALL, Ralph. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, John Wiley & Sons, 2002
- MOXON, B. Defining data mining. *DBMS, Data Warehouse Supplement*, 1998.
- FERRÃO, Francisco, *Do CRM 1.0 ao CRM 3.0*, Expresso - Caderno de Inovação e Tecnologia 2009 .
- SAS, *CRM in Banking*, A SAS White Paper, 2001

7. Anexos

7.1) Anexo 1: Descrição de variáveis da ABT

ID	NOMINAL	3
INPUT	BINARY	40
INPUT	INTERVAL	214
INPUT	NOMINAL	25
INPUT	ORDINAL	7
TARGET	BINARY	1

Tabela 7 - Tipos de variáveis da ABT

NOME	DESCRIÇÃO
ALTER_SATISF_BALCAO_ULT12M	Se o Cliente alterou o grau de satisfação pelo Balcão nos últimos 12 meses
ALTER_SATISF_BANCO_ULT12M	Se o Cliente alterou o grau de satisfação pelo Banco nos últimos 12 meses
ALTER_SATISF_CANAIS_ULT12M	Se o Cliente alterou o grau de satisfação pelos Canais nos últimos 12 meses
CCLIENTE	Código do Cliente
CCOD_DISTrito	Código de distrito do cliente
CCOD_NEGATIVOS	Código de Negativos
CCONCELHO_DESENVOL	Código associado ao Concelho em termos de Desenvolvimento
CCONCELHO_NEGOCIO	Código associado ao Concelho em termos de Negócio
CCONCELHO_RIQUEZA	Código associado ao Concelho em termos de Riqueza
CESTADO_CIVIL_CD	Código do Estado civil do cliente
CFORMA_PAGAMENTO_FREQ_MES12	Código da Forma de pagamento mais utilizada (multibanco, cheque, cartão de crédito ou cartão de débito) no mês actual
CFORMA_PAGAMENTO_FREQ_ULT12M	Código da Forma de pagamento mais utilizada (multibanco, cheque, cartão de crédito ou cartão de débito) nos últimos 12 meses

CFORMA_PAGAMENTO_FREQ_ULT3M	Código da Forma de pagamento mais utilizada (multibanco, cheque, cartão de crédito ou cartão de débito) nos últimos 3 meses
CFORMA_PAGAMENTO_FREQ_ULT6M	Código da Forma de pagamento mais utilizada (multibanco, cheque, cartão de crédito ou cartão de débito) nos últimos 6 meses
CMARCA	Código da Marca
CMES	Código do Mês
CONTAC_CALLCENTER_VALID	Indicador de contacto de call center válido
CSATISFACAO_BALCAO	Se o Cliente está Satisfeito com o Balcão
CSATISFACAO_BANCO	Cliente está satisfeito com o Banco
CSATISFACAO_CANAIS	Se o Cliente está Satisfeito com os Canais
IALTER_FORMA_PAGAM_MES12	Se o cliente alterou a forma de pagamento no mês actual
IALTER_FORMA_PAGAM_ULT12M	Se o cliente alterou a forma de pagamento nos ultimos 12 meses
IALTER_FORMA_PAGAM_ULT3M	Se o cliente alterou a forma de pagamento nos ultimos 3 meses
IALTER_FORMA_PAGAM_ULT6M	Se o cliente alterou a forma de pagamento nos ultimos 6 meses
ICCMES12	Se tem o Cartão de Crédito Activo no mês actual
ICC_ULT12M_FLG	Se o Cliente tem cartão de crédito nos ultimos 12 meses.
ICC_UTILIZADOMES12	Se tem Cartão de Crédito utilizado (dos descontos) no mês actual
ICC_UTILIZADO_ULT3M	Se tem Cartão de Crédito utilizado (dos descontos) nos ultimos 3 meses
ICC_UTILIZADO_ULT6M	Se tem Cartão de Crédito utilizado (dos descontos) nos ultimos 6 meses
ICDMES12	Se tem o Cartão de Débito Activo no mês actual
ICLIENTE_ACTIVADO_MES12	Se o cliente está activo no mês actual
ICOLABORADOR_FLG	Se o Cliente é empregado/colaborador do banco
ICOLABORADOR_REFORMADO_FLG	Se o Cliente é colaborador-reformado do Banco
ICONST_APLIC_FLG	Se o Cliente faz constituição de aplicações
IC_ESPE	Se o Cliente tem mais de 250 euros em crédito especializado.
IC_HABI	Se o cliente tem mais de 250 euros em crédito à habitação.
IC_PESSOAL	Se o Cliente tem mais de 250 euros em crédito pessoal.
IDESMOB_APLIC_FLG	Se o Cliente faz desmobilizações (cancelamento) de Aplicações/Produtos
IDCMES12	Se o Cliente tem Domiciliações de Pagamentos no mês actual
IDMC_ULT12M	Se o Cliente tem Domiciliações de Pagamentos nos ultimos 12 meses
IDMC_ULT3M	Se o Cliente tem Domiciliações de Pagamentos nos ultimos 3 meses
IDMC_ULT6M	Se o Cliente tem Domiciliações de Pagamentos nos ultimos 6 meses

IDMORMES12	Se o Cliente tem Domiciliações de Ordenados depositados à ordem no mês actual
ID_PRAZO	Se o Cliente tem mais de 250 euros em Depósitos a prazo.
IEMIGRANTE_FLG	Se o cliente é emigrante (por natureza jurídica)
IESCALAO_ANTIGMES12	Escalão de Antiguidade do Cliente no mês actual
IESCALAO_IDADEMES12	Escalão de Idade do Cliente no mês actual
IEST_COBRAN_EMPREST_CHMES12	Registo do período mais elevado de delinquência/incumprimento (em dias) que o crédito à habitação teve nos últimos 12 meses. Entram os clientes activos e os clientes não activos.
IEX_COLABORADOR_FLG	Se o Cliente é ex-empregado/colaborador do banco
IFALECIDO_FLG	Se o Cliente está falecido ou não
IFIADOR_CH	Se existe ou não fiador no crédito à habitação activo
IFUNDOS	Se o Cliente tem mais de 250 euros em Fundos.
IHABILTACOES_CD	Código Associado às Habilitações literárias do cliente
ILIQUID_CHMES12_ULT_12M_FLG	Se o cliente liquidou o Crédito à Habitação que tem e não possui um outro crédito no Banco nos últimos 12 meses
IMEDIADOR_IMOBIL	Indica se o Cliente foi angariado pelo mediador imobiliário
IMUDOU_BALCAO_ULT12M_FLG	Se o Cliente mudou de Balcão nos últimos 12 meses
IMUDOU_GESTOR_ULT12M_FLG	Se o Cliente mudou de Gestor nos últimos 12 meses
IMUDOU_SPREAD_CH_MES12_FLG	Se o cliente alterou ou não o spread nos últimos 12 meses
INACIONALIDADE_FLG	Se o Cliente é ou não cidadão nacional
IO_CREDITOS	Se o Cliente tem mais de 250 euros em outros créditos.
IPOUPANCAS	Se o Cliente tem mais de 250 euros em poupanças.
IPRIM_BANCO_MES12_FLG	Se o cliente possui este Banco como primeiro Banco (Modelo do primeiro Banco)
IP_ESTRUT	Se o Cliente tem mais de 250 euros em Produtos Estruturados.
IP_FISCAIS	Se o Cliente tem mais de 250 euros em Produtos Fiscais .
IQUADMES12	Quadrante do cliente (Vinculação do Cliente) no mês actual
IRECUSA_CH_FLG	Indica se o Cliente tem associado um processo de crédito à habitação recusado, sem nenhum crédito contratado
IRECUSA_CP_FLG	Se o Cliente tem associado um processo de crédito pessoal recusado sem outro contratado
ISALDO_DO_FLGMES12	Se o Saldo da DO é zero no mês actual
ISEXO_CD	Código do Sexo do cliente
IS_CAPIT	Se o Cliente tem mais de 250 euros em Seguros de Capitalização.
ITIPO_CLIENTE_MES12	
ITITULOS	Se o Cliente tem mais de 250 euros em títulos/apólices.

MORADA_VALID	Indicador de morada válida
NUM_CONTAC_UTEIS_CALLCENTER	Numero de contactos uteis do call center
NUM_MESES_DESDE_ALVO_CAMP	Número de meses desde que foi alvo de campanhas
NUM_RESP_POSIT_CAMP_ULT12M	Número de respostas positivas a campanhas nos últimos 12 meses
NUM_VEZES_ALVO_CAMP_ULT12M	Número de vezes que foi alvo de campanhas nos últimos 12 meses
NUTILIZ_HOMEBANKINGMES12	
NUTILIZ_HOMEBANKING_ULT12M	
NUTILIZ_HOMEBANKING_ULT3M	
NUTILIZ_HOMEBANKING_ULT6M	
N_MOV_CD_MES12	Número de movimentos de cartão de débito no mês actual
POTENCIAL_BANCARIO	
POTENCIAL_SE	Potencial Sócio-Económico do Cliente
TARGET_ABANDONO	Detectar os clientes que abandonam e os que se mantêm.
TRAN_BAL_EXTERNO	Nº de transacções que o cliente realizou em balcão externo
TRAN_BAL_INTERNA	Nº de transacções que o cliente realizou em balcão do grupo
TRAN_BT	Nº de transacções que o cliente realizou em bt
TRAN_NET	Nº de transacções que o cliente realizou em net
TRAN_SB	Nº de transacções que o cliente realizou no selfbanking
VALOR_AMORT_CHMES12	Valor (montante) da Amortização parcial que foi efectuada do Crédito à Habitação no mês actual
VAR_NPROD_ULT12M	Variação do Número de produtos que o cliente tem nos ultimos 12 meses
VAR_QUAD_ULT12M	Variação do Quadrante em que Cliente se insere nos ultimos 12 meses
VDECCC	Número de meses que o cliente decrementou o uso de Canais Complementares nos últimos 5 meses
VDECDO	Número de vezes que o cliente decrementou os movimentos de depósitos à ordem nos últimos 5 meses
VDECTB	Número de vezes que o cliente decrementou o envolvimento nos últimos 9 meses
VDERIVADA	Relação entre o Saldo Médio Actual do Cliente V.S. e Saldo Médio passado (Rácio)
VDMACT	Relação entre o volume Actual em activos v.s. o que ele teve nos últimos 9 meses
VDMREC	Relação entre o volume Actual em recursos v.s. o que ele teve nos últimos 9 meses
VD_NPROD	Relação existente entre o número de produtos máximo que o cliente teve nos ultimos 12 meses e o número de produtos que ele tem actualmente
VINCCC	Número de meses que o cliente incrementou o uso de Canais Complementares nos últimos 5 meses
VINCDO	Número de vezes que o cliente incrementou os movimentos de depósitos à ordem nos últimos 5 meses
VINCTB	Número de vezes que o cliente incrementou o envolvimento nos últimos 9 meses

VMANCC	Número de meses que o cliente manteve o uso de Canais Complementares nos últimos 5 meses
VMANDO	Número de vezes que o cliente manteve o Montante de depósitos à ordem nos últimos 5 meses
VMANTB	Número de vezes que o cliente manteve o envolvimento nos últimos 9 meses
VMAX_MESES_ULT_ACESSO_BT_ULT12M	Número máximo de meses sem acesso ao serviço de Banca Telefónica nos últimos 12 meses
VMONTB	Relação entre o montante de envolvimento que o cliente tem no momento v.s. o máximo que já deteve
VMON_ACTMES12	Montante de Activos que o cliente tem no mês actual
VMON_AMORT_CREDMES12	Montante de Amortizações de Créditos feitas no mês actual
VMON_AMORT_CRED_ULT3M	Montante de Amortizações de Créditos feitas nos últimos 3 meses
VMON_AMORT_CRED_ULT6M	Montante de Amortizações de Créditos feitas nos últimos 6 meses
VMON_CHEQUES_OUTMES12	Montante de Cheques Out no mês actual
VMON_CHEQ_OUT_ULT12M	Montante de Cheques Out nos últimos 12 meses
VMON_CHEQ_OUT_ULT3M	Montante de Cheques Out nos últimos 3 meses
VMON_CHEQ_OUT_ULT6M	Montante de Cheques Out nos últimos 6 meses
VMON_CHMES12	Montante (volume) que o cliente detém em Crédito à Habitação no mês actual
VMON_DEPOSITOS_DOMES12	Montante de Depósitos feitos pelo cliente no mês actual
VMON_DEPOSITOS_DO_ULT12M	Montante de Depósitos feitos pelo cliente nos últimos 12 meses
VMON_DEPOSITOS_DO_ULT3M	Montante de Depósitos feitos pelo cliente nos últimos 3 meses
VMON_DEPOSITOS_DO_ULT6M	Montante de Depósitos feitos pelo cliente nos últimos 6 meses
VMON_DISPONIB_CREDMES12	Montante de Disponibilização de Créditos no mês actual
VMON_DISPONIB_CRED_ULT12M	Montante de Disponibilização de Créditos nos últimos 12 meses
VMON_DISPONIB_CRED_ULT3M	Montante de Disponibilização de Créditos nos últimos 3 meses
VMON_DISPONIB_CRED_ULT6M	Montante de Disponibilização de Créditos nos últimos 6 meses
VMON_DIVIDA	Montante em dívida que o cliente tem no momento em crédito à habitação
VMON_FALTA_PAGAR_CEMES12	Montante a pagar crédito especializado até ao termino do contrato no mês actual
VMON_FALTA_PAGAR_CPMES12	Montante a pagar de crédito pessoal até ao termino do contrato no mês actual
VMON_FALTA_PAGAR_OCMES12	Montante a pagar de outros créditos até ao termino do contrato no mês actual
VMON_GASTOS_TOTAIS_ULT12M	Montante de Gastos Totais nos últimos 12 meses
VMON_GASTOS_TOTAIS_ULT3M	Montante de Gastos Totais nos últimos 3 meses
VMON_GASTOS_TOTAIS_ULT6M	Montante de Gastos Totais nos últimos 6 meses
VMON_LEVANT_DOMES12	Montante de Levantamentos feitos pelo cliente no mês actual
VMON_LEVANT_DO_ULT12M	Montante de Levantamentos feitos pelo cliente nos últimos 12 meses
VMON_LEVANT_DO_ULT3M	Montante de Levantamentos feitos pelo cliente nos últimos 3 meses

VMON_LEVANT_DO_ULT6M	Montante de Levantamentos feitos pelo cliente nos últimos 6 meses
VMON_LEVANT_NACI_CDMES12	Montante de Levantamentos Nacionais feitos no mês actual
VMON_LEVANT_TOT_CD_ULT6M	Montante de Levantamentos Totais feitos nos últimos 6 meses
VMON_MAX_SALDO_PE_ULT12M	Saldo (montante) máximo da conta de produtos estruturados que o cliente tem nos ultimos 12 meses
VMON_MAX_SALDO_T_ULT12M	Saldo (montante) máximo da conta de títulos que o cliente tem nos ultimos 12 meses
VMON_MOV_CRED_DOMES12	Montante de Movimentos a crédito dos depósitos à ordem no mês actual
VMON_MOV_DEBIT_DOMES12	Montante de Movimentos a débito dos depósitos à ordem no mês actual
VMON_O_CREDITOS_DOMES12	Montante de Outros créditos (DO) feitos no mês actual
VMON_O_CREDITOS_DO_ULT12M	Montante de Outros créditos (DO) feitos nos ultimos 12 meses
VMON_O_CREDITOS_DO_ULT3M	Montante de Outros créditos (DO) feitos nos ultimos 3 meses
VMON_O_CREDITOS_DO_ULT6M	Montante de Outros créditos (DO) feitos nos ultimos 6 meses
VMON_O_DEBITOS_DOMES12	Montante de Outros débitos (DO) feitos no mês actual
VMON_O_DEBITOS_DO_ULT12M	Montante de Outros débitos (DO) feitos nos ultimos 12 meses
VMON_O_DEBITOS_DO_ULT3M	Montante de Outros débitos (DO) feitos nos ultimos 3 meses
VMON_O_DEBITOS_DO_ULT6M	Montante de Outros débitos (DO) feitos nos ultimos 6 meses
VMON_PAGAM_PRESTMES12	Montante de Pagamentos de Prestações feitas pelo cliente no mês actual
VMON_PAGAM_PREST_ULT12M	Montante de Pagamentos de Prestações feitas pelo cliente nos ultimos 12 meses
VMON_PAGAM_PREST_ULT3M	Montante de Pagamentos de Prestações feitas pelo cliente nos ultimos 3 meses
VMON_PAGAM_PREST_ULT6M	Montante de Pagamentos de Prestações feitas pelo cliente nos ultimos 6 meses
VMON_PAGO_CE_ULT6M	Montante Pago do Crédito Especializado nos ultimos 6 meses
VMON_PAGO_CP_MES12	Montante do Crédito Pessoal pago no mês actual
VMON_PAGO_OC_ULT12M	Montante de outros créditos pago nos ultimos 12 meses
VMON_PAGO_OC_ULT3M	Montante de outros créditos pago nos ultimos 3 meses
VMON_PAGO_OC_ULT6M	Montante de outros créditos pago nos ultimos 6 meses
VMON_PAGO_OC_ULTMES12	Montante de outros créditos pago no mês actual
VMON_PROD_BANCARIOMES12	Montante do Produto Bancário da MIT90 (anualizado) no mês actual
VMON_PROD_BANCARIO_12M	Montante do Produto Bancário da MIT90 (anualizado) nos ultimos 12 meses
VMON_PROD_BANCARIO_3M	Montante do Produto Bancário da MIT90 (anualizado) nos ultimos 3 meses
VMON_PROD_BANCARIO_6M	Montante do Produto Bancário da MIT90 (anualizado) nos ultimos 6 meses
VMON_RECME12	Montante de Recursos que o cliente tem no mês actual
VMON_SALDO_DPMES12	Saldo (montante) da conta de depósitos a prazo que o cliente tem no mês actual
VMON_SALDO_DP_UTL3M	

VMON_SALDO_DP_UTL6M	
VMON_SALDO_FUMES12	Saldo (montante) da conta de fundos que o cliente tem no mês actual
VMON_SALDO_FU_ULT3M	Saldo (montante) da conta de fundos que o cliente tem nos ultimos 3 meses
VMON_SALDO_FU_ULT6M	Saldo (montante) da conta de fundos que o cliente tem nos ultimos 6 meses
VMON_SALDO_PEMES12	Saldo (montante) da conta de produtos estruturados que o cliente tem no mês actual
VMON_SALDO_PE_ULT3M	Saldo (montante) da conta de produtos estruturados que o cliente tem nos ultimos 3 meses
VMON_SALDO_PE_ULT6M	Saldo (montante) da conta de produtos estruturados que o cliente tem nos ultimos 6 meses
VMON_SALDO_PFIMES12	Saldo (montante) da conta de produtos fiscais que o cliente tem no mês actual
VMON_SALDO_PFI_ULT3M	Saldo (montante) da conta de produtos fiscais que o cliente tem nos ultimos 3 meses
VMON_SALDO_PFI_ULT6M	Saldo (montante) da conta de produtos fiscais que o cliente tem nos ultimos 6 meses
VMON_SALDO_POMES12	Saldo (montante) da conta de poupanças que o cliente tem no mês actual
VMON_SALDO_PO_ULT6M	Saldo (montante) da conta de poupanças que o cliente tem nos ultimos 6 meses
VMON_SALDO_SEMES12	Saldo (montante) da conta de seguros de capitalização que o cliente tem no mês actual
VMON_SALDO_SE_ULT3M	Saldo (montante) da conta de seguros de capitalização que o cliente tem nos ultimos 3 meses
VMON_SALDO_SE_ULT6M	Saldo (montante) da conta de seguros de capitalização que o cliente tem nos ultimos 6 meses
VMON_SALDO_TMES12	Saldo (montante) da conta de títulos que o cliente tem no mês actual
VMON_SALDO_T_ULT12M	Saldo (montante) da conta de títulos que o cliente tem nos ultimos 12 meses
VMON_SALDO_T_ULT3M	Saldo (montante) da conta de títulos que o cliente tem nos ultimos 3 meses
VMON_SALDO_T_ULT6M	Saldo (montante) da conta de títulos que o cliente tem nos ultimos 6 meses
VMON_SMCVMES12	Montante do Saldo Médio Credito Vencido dos depósitos à ordem no mês actual
VMON_SMDNAMES12	Montante do Saldo Médio Descoberto não Autorizado dos depósitos à ordem no mês actual
VMON_SMDOAMES12	Montante do Saldo Médio Descoberto Autorizado dos depósitos à ordem no mês actual
VMON_SMDOMES12	Montante do Saldo Médio dos depósitos à ordem no mês actual
VMON_TITULOSMES12	Montante de titulos que o cliente tem no mês actual
VMON_TRANS_BANCMES12	Montante de Transferências efectuadas no mês actual
VMON_TRANS_BANC_ULT12M	Montante de Transferências efectuadas nos últimos 12 meses
VMON_TRANS_BANC_ULT3M	Montante de Transferências efectuadas nos últimos 3 meses
VMON_TRANS_BANC_ULT6M	Montante de Transferências efectuadas nos últimos 6 meses
VMON_TRANS_INTER_BANCMES12	Montante de Transferências Inter-Bancárias efectuadas no mês actual
VMON_TRANS_INTER_BANC_ULT12M	Montante de Transferências Inter-Bancárias efectuadas nos últimos 12 meses
VMON_TRANS_INTER_BANC_ULT3M	Montante de Transferências Inter-Bancárias efectuadas nos últimos 3 meses
VMON_TRANS_INTER_BANC_ULT6M	Montante de Transferências Inter-Bancárias efectuadas nos últimos 6 meses

VMON_TRANS_OUTMES12	Montante de Transferências Out no mês actual
VMON_TRANS_OUT_ULT12M	Montante de Transferências Out nos ultimos 12 meses
VMON_TRANS_OUT_ULT3M	Montante de Transferências Out nos ultimos 3 meses
VMON_TRANS_OUT_ULT6M	Montante de Transferências Out nos ultimos 6 meses
VM_FALTA_CESS_CRE_HAB	Indicação do tempo (Número de meses) que falta para o término do Crédito à Habitação da última conta a vencer (maior data de amortização dos créditos activos)
VNPRODME12	Número de Produtos que o cliente tem no mês actual
VNULCC	Número de meses que o cliente não utilizou Canais Complementares nos últimos 5 meses
VNULDO	Número de vezes que o cliente não detinha montante de depósitos à ordem nos últimos 5 meses
VNULTB	Número de vezes que o cliente não apresentava volume de negócio nos últimos 9 meses
VNUMM_FALTA_CESS_CP	Número de Indicação do tempo que falta para o término do Crédito Pessoal (ultima conta do crédito)
VNUMM_ORDENADO	Número de Meses de ordenado
VNUM_AMORT_CH_ULT12M	Número de Amortizações parciais nos ultimos 12 meses
VNUM_AMORT_CREDME12	Número de Amortizações de Créditos feitas no mês actual
VNUM_AMORT_CRED_ULT3M	Número de Amortizações de Créditos feitas nos ultimos 3 meses
VNUM_AMORT_CRED_ULT6M	Número de Amortizações de Créditos feitas nos ultimos 6 meses
VNUM_ANULACOES_CCMES12	Número de Anulações feitas no mês actual
VNUM_CAMBIOS	Número de Movimentos de Câmbios
VNUM_CHEQUES_OUTMES12	Número de Cheques Out no mês actual
VNUM_CHEQ_OUT_ULT12M	Número de Cheques Out nos ultimos 12 meses
VNUM_CHEQ_OUT_ULT3M	Número de Cheques Out nos ultimos 3 meses
VNUM_CHEQ_OUT_ULT6M	Número de Cheques Out nos ultimos 6 meses
VNUM_COMP_INTERNETMES12	Número de Compras feitas pela internet no mês actual
VNUM_COMP_INTERNET_ULT12M	Número de Compras feitas pela internet nos últimos 12 meses
VNUM_COMP_INTERNET_ULT3M	Número de Compras feitas pela internet nos últimos 3 meses
VNUM_COMP_INTERNET_ULT6M	Número de Compras feitas pela internet nos últimos 6 meses
VNUM_CTA_CRED_PESS_FECHADAS	Número de Contas de crédito pessoal fechadas em que o cliente é primeiro titular
VNUM_DEPOSITOS_DOMES12	Número de Depósitos feitos pelo cliente no mês actual
VNUM_DEPOSITOS_DO_ULT12M	Número de Depósitos feitos pelo cliente nos ultimos 12 meses
VNUM_DEPOSITOS_DO_ULT3M	Número de Depósitos feitos pelo cliente nos ultimos 3 meses
VNUM_DEPOSITOS_DO_ULT6M	Número de Depósitos feitos pelo cliente nos ultimos 6 meses
VNUM_DIAS_ULT_AMORT_CH	Número de dias desde a ultima amortização parcial

VNUM_DISPONIB_CREDMES12	Numero de Disponibilizações de Créditos no mês actual
VNUM_DISPONIB_CRED_ULT3M	Numero de Disponibilizações de Créditos nos últimos 3 meses
VNUM_DISPONIB_CRED_ULT6M	Numero de Disponibilizações de Créditos nos últimos 6 meses
VNUM_GASTOS_TOTAIS_MES12	Numero de Gastos Totais no mes actual
VNUM_GASTOS_TOTAIS_ULT12M	Numero de Gastos Totais nos ultimos 12 meses
VNUM_GASTOS_TOTAIS_ULT3M	Numero de Gastos Totais nos ultimos 3 meses
VNUM_GASTOS_TOTAIS_ULT6M	Numero de Gastos Totais nos ultimos 6 meses
VNUM_LEVANT_DOMES12	Número de Levantamentos feitos pelo cliente no mês actual
VNUM_LEVANT_DO_ULT12M	Número de Levantamentos feitos pelo cliente nos últimos 12 meses
VNUM_LEVANT_DO_ULT3M	Número de Levantamentos feitos pelo cliente nos últimos 3 meses
VNUM_LEVANT_DO_ULT6M	Número de Levantamentos feitos pelo cliente nos últimos 6 meses
VNUM_MESES_FINANCIA_CP	Prazo máximo de Financiamento do Crédito Pessoal
VNUM_MESES_ULT_QUEST_SATISF	Numero de Meses desde o ultimo questionario sobre o nível de satisfação do Cliente
VNUM_MOVMES12	Movimentos que o cliente realizou (depósitos à ordem + cartão de crédito) no mês actual
VNUM_MOV_CRED_DOMES12	Número de Movimentos a crédito dos depósitos à ordem no mês actual
VNUM_MOV_DEBIT_DOMES12	Número de Movimentos a débito dos depósitos à ordem no mês actual
VNUM_O_CREDITOS_DOMES12	Número de Outros créditos (DO) feitos no mês actual
VNUM_O_CREDITOS_DO_ULT12M	Número de Outros créditos (DO) feitos nos ultimos 12 meses
VNUM_O_CREDITOS_DO_ULT3M	Número de Outros créditos (DO) feitos nos ultimos 3 meses
VNUM_O_CREDITOS_DO_ULT6M	Número de Outros créditos (DO) feitos nos ultimos 6 meses
VNUM_O_DEBITOS_DOMES12	Número de Outros débitos (DO) feitos no mês actual
VNUM_O_DEBITOS_DO_ULT12M	Número de Outros débitos (DO) feitos nos ultimos 12 meses
VNUM_O_DEBITOS_DO_ULT3M	Número de Outros débitos (DO) feitos nos ultimos 3 meses
VNUM_O_DEBITOS_DO_ULT6M	Número de Outros débitos (DO) feitos nos ultimos 6 meses
VNUM_PAGAM_PRESTMES12	Número de Pagamentos de Prestações feitas pelo cliente no mês actual
VNUM_PAGAM_PREST_ULT12M	Número de Pagamentos de Prestações feitas pelo cliente nos ultimos 12 meses
VNUM_PAGAM_PREST_ULT3M	Número de Pagamentos de Prestações feitas pelo cliente nos ultimos 3 meses
VNUM_PAGAM_PREST_ULT6M	Número de Pagamentos de Prestações feitas pelo cliente nos ultimos 6 meses
VNUM_TRANS_BANCMES12	Número de Transferências efectuadas no mês actual
VNUM_TRANS_BANC_ULT12M	Número de Transferências efectuadas nos últimos 12 meses
VNUM_TRANS_BANC_ULT3M	Número de Transferências efectuadas nos últimos 3 meses
VNUM_TRANS_BANC_ULT6M	Número de Transferências efectuadas nos últimos 6 meses

VNUM_TRANS_INTER_BANCMES12	Número de Transferências Inter-Bancárias efectuadas no mês actual
VNUM_TRANS_INTER_BANC_ULT3M	Número de Transferências Inter-Bancárias efectuadas nos últimos 3 meses
VNUM_TRANS_INTER_BANC_ULT6M	Número de Transferências Inter-Bancárias efectuadas nos últimos 6 meses
VNUM_TRANS_OUTMES12	Número de Transferências Out no mês actual
VNUM_TRANS_OUT_ULT12M	Número de Transferências Out nos ultimos 12 meses
VNUM_TRANS_OUT_ULT3M	Número de Transferências Out nos ultimos 3 meses
VNUM_TRANS_OUT_ULT6M	Número de Transferências Out nos ultimos 6 meses
VOL_NEGMES12	Volume de Negócios, excepto crédito à habitação (Recursos+Activo+Títulos - Crédito à Habitação) no mês actual
VPAGO_CH_MES12	
VPAGO_CH_ULT3M	Montante Pago do Crédito à Habitação pago nos ultimos 3 meses (parcela paga neste mês)
VPCT_APLICADO_PROD_RISCO_ULT3M	Montante aplicado em Produtos de Risco (Produtos Estruturados, Produtos fiscais, Fundos) no mês actual face ao montante gasto nos ultimos 3 meses
VPCT_APLICADO_PROD_RISCO_ULT6M	Montante aplicado em Produtos de Risco (Produtos Estruturados, Produtos fiscais, Fundos) no mês actual face ao montante gasto nos ultimos 6 meses
VPERIODO_DESMOB_APLIC	Indicador se o cliente fez desmobilização no periodo
VPERIODO_ULTMOV_DO	Número de dias desde o ultimo movimento da conta à ordem
VPERIOD_CONST_APLIC	Período desde a desmobilização de aplicações
VP_ACT_CE	Montante que o cliente tem em activos excepto crédito especializado no mês actual
VP_ACT_CH	Montante que o cliente tem em activos excepto crédito à habitação
VP_ACT_CP	Montante que o cliente tem em activos excepto crédito pessoal
VP_ACT_OC	Montante que o cliente tem em activos excepto outros créditos
VRENDIM_MEDIO	Montante do Rendimento Médio do Cliente
VSPREAD_ACTUAL_CH_MES12	
VTOT_RECLAMAMES12	Número total de Reclamações do Cliente no mês actual
VTOT_RECLAMA_ULT3M	
VTOT_RECLAMA_ULT6M	Número total de Reclamações do Cliente nos últimos 6 meses
VUTILIZ_BANCA_TELEFMES12	Número de utilizações (log in) de Banca Telefónica por parte do utilizador no mês actual
VUTILIZ_BANCA_TELEF_ULT12M	Número de utilizações (log in) de Banca Telefónica por parte do utilizador nos ultimos 12 meses
VUTILIZ_BANCA_TELEF_ULT3M	Número de utilizações (log in) de Banca Telefónica por parte do utilizador nos ultimos 3 meses
VUTILIZ_BANCA_TELEF_ULT6M	Número de utilizações (log in) de Banca Telefónica por parte do utilizador nos ultimos 6 meses

Tabela 8 – Descrição de variáveis da ABT

7.2) Anexo 2: Distribuição de valores das variáveis sócio-demográficas

V_TARGET			
VALOR	SIGNIFICADO	FREQUENCY	PERCENT
0	Não Abandono	964679	99,17%
1	Abandono	8027	0,83%

Tabela 9 – Distribuição de valores da variável target

IESCALAO_IDADEMES12 (Idade do cliente no mês actual)			
VALOR	SIGNIFICADO	FREQUENCY	PERCENT
3	Entre 18-24 Anos	32294	3,32%
4	Entre 25-29 Anos	64782	6,66%
5	Entre 30-39 Anos	258059	26,53%
6	Entre 40-54 Anos	356886	36,69%
7	Entre 55-64 Anos	146490	15,06%
8	Mais de 65 Anos	114196	11,74%

Tabela 10 - Distribuição de clientes por escalão etário

ISEXO_CD (Codigo do Sexo do cliente)			
VALOR	SIGNIFICADO	FREQUENCY	PERCENT
F	Feminino	369045	37,94%
M	Masculino	603661	62,06%

Tabela 11 – Distribuição de clientes por género

CESTADO_CIVIL_CD (Estado Civil do cliente no mês actual)			
VALOR	SIGNIFICADO	FREQUENCY	PERCENT
1	Solteiro	221485	22,77%
2	Casado / União Facto	637706	65,56%
3	Divorciado	81318	8,36%
4	Viúvo	30932	3,18%
5	Não Catalogado	1265	0,13%

Tabela 12 – Distribuição de clientes por Estado Civil

INACIONALIDADE_FLG (Se o Cliente é ou não cidadão nacional)			
VALOR	SIGNIFICADO	FREQUENCY	PERCENT
0	Estrangeiro	21594	2,22%
1	Português	951112	97,78%

Tabela 13 – Distribuição de clientes por nacionalidade

IHABILTACOES_CD (Habilitações literárias do cliente no mês actual)			
VALOR	SIGNIFICADO	FREQUENCY	PERCENT
1	Licenciatura	330234	33,95%
2	Mestrado	53402	5,49%
3	Frequência Universitária	128203	13,18%
4	12º Ano ou inferior	460868	47,38%

Tabela 14 - Distribuição de clientes por Habilitações Literárias

POTENCIAL_BANCARIO (Potencial bancário associado ao Cliente)			
VALOR	SIGNIFICADO	FREQUENCY	PERCENT
1	Muito Elevado	7782	0,80%
2	Elevado	89003	9,15%
3	Médio	278972	28,68%
4	Baixo	542867	55,81%
5	Muito Baixo ou Nulo	54082	5,56%

Tabela 15 - Distribuição de clientes por Potencial Bancário

POTENCIAL_SE (Potencial Sócio Económico associado ao Cliente)			
VALOR	SIGNIFICADO	FREQUENCY	PERCENT
1	Muito Elevado	35406	3,64%
2	Elevado	139875	14,38%
3	Médio	431492	44,36%
4	Baixo	196195	20,17%
5	Não Catalogado	169737	17,45%

Tabela 16 - Distribuição de clientes por Potencial Sócio Económico

CCONCELHO_DESENVOL (Desenvolvimento associado ao Concelho/Zona Habitacional)			
VALOR	SIGNIFICADO	FREQUENCY	PERCENT
1	Metropolitano	421960	43,38%
2	Urbano	356205	36,62%
3	Rural	161858	16,64%
4	Aldeia	12256	1,26%
5	Estrangeiro	15369	1,58%
6	Não Catalogado	5058	0,52%

Tabela 17 - Distribuição de clientes por tipo de concelho (desenvolvimento)

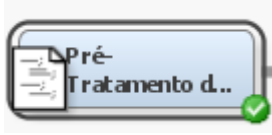

CCONCELHO_RIQUEZA (Riqueza associada ao Concelho/Zona Habitacional)			
VALOR	SIGNIFICADO	FREQUENCY	PERCENT
1	Rico	187246	19,25%
2	Médio Alto	277610	28,54%
3	Médio	440441	45,28%
4	Pobre	62350	6,41%
5	Não Catalogado	5058	0,52%

Tabela 18 - Distribuição de clientes por tipo de concelho (riqueza)

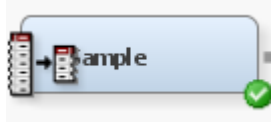
CCONCELHO_NEGOCIO (Negócio associado ao Concelho/Zona Habitacional)			
VALOR	SIGNIFICADO	FREQUENCY	PERCENT
1	Recursos	218762	22,49%
2	Iminentemente Recursos	108554	11,16%
3	Iminentemente Crédito	430033	44,21%
4	Crédito	210299	21,62%
5	Não Catalogado	5058	0,52%

Tabela 19 - Distribuição de clientes por tipo de concelho (negócio)


7.3) Anexo 3: SAS Code utilizado

	<pre>data &EM_EXPORT_TRAIN;set &EM_IMPORT_DATA;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;</pre>
	<pre>libname tfc 'C:\ABT Abandono Cliente'; data tfc.result_modelo; set &EM_IMPORT_SCORE; run;</pre>

7.4) Anexo 4: Node Sample


	<table><tr><th>Property</th><th>Value</th></tr><tr><td>Size</td><td></td></tr><tr><td>Type</td><td>Percentage</td></tr><tr><td>Observations</td><td>1</td></tr><tr><td>Percentage</td><td>100.0</td></tr><tr><td>Alpha</td><td>0.01</td></tr><tr><td>PValue</td><td>0.01</td></tr><tr><td>Cluster Method</td><td>Random</td></tr><tr><td>Stratified</td><td></td></tr><tr><td>Criterion</td><td>Level Based</td></tr><tr><td>Ignore Small Strata</td><td>No</td></tr><tr><td>Minimum Strata Size</td><td>5</td></tr><tr><td>Level Based Options</td><td></td></tr><tr><td>Level Selection</td><td>Event</td></tr><tr><td>Level Proportion</td><td>100.0</td></tr><tr><td>Sample Proportion</td><td>50.0</td></tr></table>	Property	Value	Size		Type	Percentage	Observations	1	Percentage	100.0	Alpha	0.01	PValue	0.01	Cluster Method	Random	Stratified		Criterion	Level Based	Ignore Small Strata	No	Minimum Strata Size	5	Level Based Options		Level Selection	Event	Level Proportion	100.0	Sample Proportion	50.0
Property	Value																																
Size																																	
Type	Percentage																																
Observations	1																																
Percentage	100.0																																
Alpha	0.01																																
PValue	0.01																																
Cluster Method	Random																																
Stratified																																	
Criterion	Level Based																																
Ignore Small Strata	No																																
Minimum Strata Size	5																																
Level Based Options																																	
Level Selection	Event																																
Level Proportion	100.0																																
Sample Proportion	50.0																																

7.5) Anexo 5: Node Data Partition



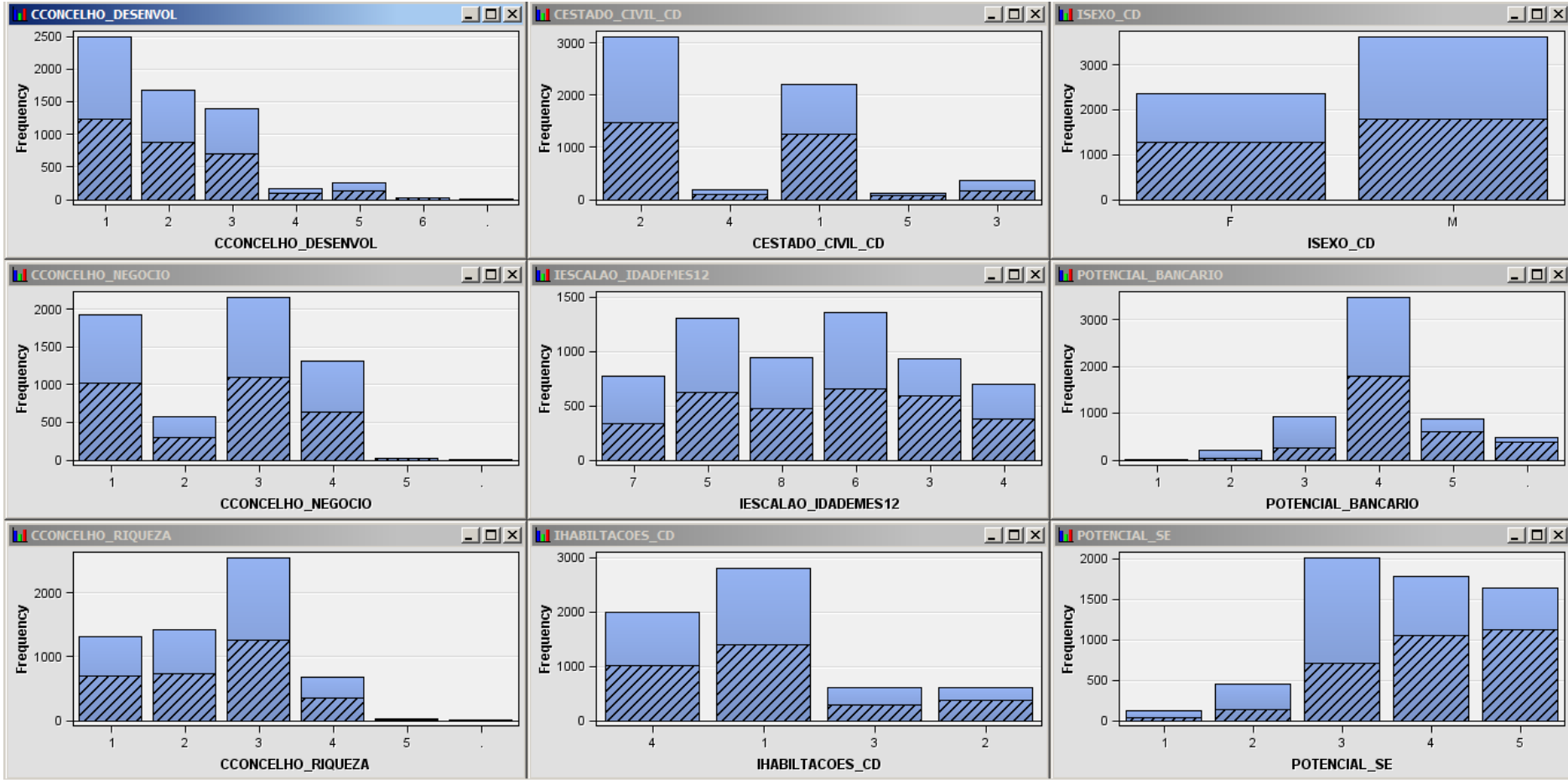
Property	Value
General	
Node ID	Part
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Output Type	Data
Partitioning Method	Stratified
Random Seed	12345
Data Set Allocations	
Training	70.0
Validation	30.0
Test	0.0

7.6) Anexo 6: Node Impute



Property	Value
General	
Node ID	Impt
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Non Missing Variables	No
Missing Cutoff	50.0
Class Variables	
Default Input Method	Tree
Default Target Method	None
Normalize Values	Yes
Interval Variables	
Default Input Method	Tree
Default Target Method	None
Default Constant Value	
Default Character Value	D
Default Number Value	0.0
Method Options	
Random Seed	12345
Tuning Parameters	...
Tree Imputation	...


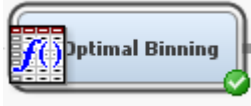
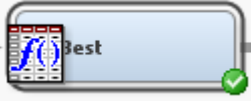
7.7) Anexo 7: Relacionamento da variável Target com as Sócio-Demográficas.



7.8) Anexo 8: Node Filter

Filter	<table> <tr> <th>Property</th><th>Value</th></tr> <tr> <td>Class Variables</td><td></td></tr> <tr> <td>Class Variables</td><td>...</td></tr> <tr> <td>Default Filtering Method</td><td>Rare Values (Percentage)</td></tr> <tr> <td>Keep Missing Values</td><td>Yes</td></tr> <tr> <td>Normalized Values</td><td>Yes</td></tr> <tr> <td>Minimum Frequency Cutoff</td><td>1</td></tr> <tr> <td>Minimum Cutoff for Percentage</td><td>0.01</td></tr> <tr> <td>Maximum Number of Levels Cutoff</td><td>25</td></tr> <tr> <td>Interval Variables</td><td></td></tr> <tr> <td>Interval Variables</td><td>...</td></tr> <tr> <td>Default Filtering Method</td><td>Extreme Percentiles</td></tr> <tr> <td>Keep Missing Values</td><td>Yes</td></tr> <tr> <td>Tuning Parameters</td><td>...</td></tr> </table>	Property	Value	Class Variables		Class Variables	...	Default Filtering Method	Rare Values (Percentage)	Keep Missing Values	Yes	Normalized Values	Yes	Minimum Frequency Cutoff	1	Minimum Cutoff for Percentage	0.01	Maximum Number of Levels Cutoff	25	Interval Variables		Interval Variables	...	Default Filtering Method	Extreme Percentiles	Keep Missing Values	Yes	Tuning Parameters	...
Property	Value																												
Class Variables																													
Class Variables	...																												
Default Filtering Method	Rare Values (Percentage)																												
Keep Missing Values	Yes																												
Normalized Values	Yes																												
Minimum Frequency Cutoff	1																												
Minimum Cutoff for Percentage	0.01																												
Maximum Number of Levels Cutoff	25																												
Interval Variables																													
Interval Variables	...																												
Default Filtering Method	Extreme Percentiles																												
Keep Missing Values	Yes																												
Tuning Parameters	...																												

7.9) Anexo 9: Transformation Nodes

	<table> <tr> <th>Property</th><th>Value</th></tr> <tr> <td>Default Methods</td><td></td></tr> <tr> <td>Interval Inputs</td><td>Log 10</td></tr> <tr> <td>Interval Targets</td><td>None</td></tr> <tr> <td>Class Inputs</td><td>Group rare levels</td></tr> <tr> <td>Class Targets</td><td>None</td></tr> <tr> <td>Treat Missing as Level</td><td>No</td></tr> </table>	Property	Value	Default Methods		Interval Inputs	Log 10	Interval Targets	None	Class Inputs	Group rare levels	Class Targets	None	Treat Missing as Level	No
Property	Value														
Default Methods															
Interval Inputs	Log 10														
Interval Targets	None														
Class Inputs	Group rare levels														
Class Targets	None														
Treat Missing as Level	No														
	<table> <tr> <th>Property</th><th>Value</th></tr> <tr> <td>Default Methods</td><td></td></tr> <tr> <td>Interval Inputs</td><td>Optimal Binning</td></tr> <tr> <td>Interval Targets</td><td>None</td></tr> <tr> <td>Class Inputs</td><td>Group rare levels</td></tr> <tr> <td>Class Targets</td><td>None</td></tr> <tr> <td>Treat Missing as Level</td><td>No</td></tr> </table>	Property	Value	Default Methods		Interval Inputs	Optimal Binning	Interval Targets	None	Class Inputs	Group rare levels	Class Targets	None	Treat Missing as Level	No
Property	Value														
Default Methods															
Interval Inputs	Optimal Binning														
Interval Targets	None														
Class Inputs	Group rare levels														
Class Targets	None														
Treat Missing as Level	No														
	<table> <tr> <th>Property</th><th>Value</th></tr> <tr> <td>Default Methods</td><td></td></tr> <tr> <td>Interval Inputs</td><td>Best</td></tr> <tr> <td>Interval Targets</td><td>None</td></tr> <tr> <td>Class Inputs</td><td>Group rare levels</td></tr> <tr> <td>Class Targets</td><td>None</td></tr> <tr> <td>Treat Missing as Level</td><td>No</td></tr> </table>	Property	Value	Default Methods		Interval Inputs	Best	Interval Targets	None	Class Inputs	Group rare levels	Class Targets	None	Treat Missing as Level	No
Property	Value														
Default Methods															
Interval Inputs	Best														
Interval Targets	None														
Class Inputs	Group rare levels														
Class Targets	None														
Treat Missing as Level	No														

7.10) Anexo 10: Node Variable Selection

Property	Value
Train	
Variables	
Max Class Level	100
Max Missing Percentage	50
Target Model	R and Chi Square

7.11) Anexo 11: Ranking de importância das variáveis

Variable Name	Chi-Square	Pr > ChiSq	Descrição da variável
OPT_IMP_VNUM_MOVMES12	2.941.526	<.0001	Movimentos que o cliente realizou (depósitos à ordem + cartão de crédito) no mês actual
IDMORMES12	1.218.352	<.0001	Se o Cliente tem Domiciliações de Ordenados depositados à ordem no mês actual
OPT_VDMREC	1.194.124	<.0001	Relação entre o volume Actual em recursos v.s. o que ele teve nos últimos 9 meses
OPT_VM_FALTA_CESS_CRE_HAB	1.169.638	<.0001	Indicação do tempo (Número de meses) que falta para o término do Crédito à Habitação da última conta a vencer (maior data de amortização dos créditos activos)
OPT_VMON_SALDO_DPMES12	828.932	<.0001	Saldo (montante) da conta de depósitos a prazo que o cliente tem no mês actual
OPT_VMON_RECMES12	781.428	<.0001	Montante de Recursos que o cliente tem no mês actual
OPT_VMON_SALDO_PFIMES12	767.746	<.0001	Saldo (montante) da conta de produtos fiscais que o cliente tem no mês actual
IDMCMES12	757.449	<.0001	Se o Cliente tem Domiciliações de Pagamentos no mês actual
OPT_VD_NPROD	728.473	<.0001	Relação existente entre o número de produtos máximo que o cliente teve nos últimos 12 meses e o número de produtos que ele tem actualmente
IFUNDOS	710.196	<.0001	Se o Cliente tem mais de 250 euros em Fundos.
OPT_IMP_VNUM_MOV_CRED_DOMES12	651.256	<.0001	Número de Movimentos a crédito dos depósitos à ordem no mês actual
OPT_VMON_SALDO_SEMES12	616.672	<.0001	Saldo (montante) da conta de seguros de capitalização que o cliente tem no mês actual
OPT_IMP_VDERIVADA	525.689	<.0001	Relação entre o Saldo Médio Actual do Cliente V.S. e Saldo Médio passado (Rácio)
OPT_IMP_VPERIODO_ULTMOV_DO	512.827	<.0001	Número de dias desde o último movimento da conta à ordem

LG10_IMP_VMONTB	509.114	<.0001	Relação entre o montante de envolvimento que o cliente tem no momento v.s. o máximo que já deteve
OPT_VMON_SMCVMES12	504.894	<.0001	Montante do Saldo Médio Credito Vencido dos depósitos à ordem no mês actual
OPT_IMP_VMON_SALDO_T_ULT6M	503.850	<.0001	Saldo (montante) da conta de títulos que o cliente tem nos últimos 6 meses
OPT_VMON_SALDO_FUMES12	477.653	<.0001	Saldo (montante) da conta de fundos que o cliente tem no mês actual
OPT_VMON_SMDOAMES12	411.323	<.0001	Montante do Saldo Médio Descoberto Autorizado dos depósitos à ordem no mês actual
OPT_VMON_SALDO_POMES12	400.158	<.0001	Saldo (montante) da conta de poupanças que o cliente tem no mês actual

Tabela 20– Ranking de importância das variáveis

7.12) Anexo 12: Assessment Score Ranking

Percentile	Lift	Cumulative Lift	% Response	Cumulative % Response	Number of Observations	Mean Posterior Probability
5	1,97510	1,97510	98,7552%	98,7552%	241	0.98913
10	1,96680	1,97095	98,3402%	98,5477%	241	0.97503
15	1,91701	1,95297	95,8506%	97,6487%	241	0.96187
20	1,88382	1,93568	94,1909%	96,7842%	241	0.94119
25	1,82573	1,91369	91,2863%	95,6846%	241	0.91107
30	1,67635	1,87414	83,8174%	93,7068%	241	0.87055
35	1,52697	1,82454	76,3485%	91,2270%	241	0.81067
40	1,59336	1,79564	79,6680%	89,7822%	241	0.74364
45	1,36929	1,74827	68,4647%	87,4136%	241	0.67515
50	1,19167	1,69282	59,5833%	84,6409%	240	0.60183
55	1,13693	1,64226	56,8465%	82,1132%	241	0.52650
60	0,76349	1,56901	38,1743%	78,4504%	241	0.40149
65	0,61411	1,49553	30,7054%	74,7765%	241	0.27251
70	0,34855	1,41358	17,4274%	70,6789%	241	0.14593
75	0,13278	1,32817	6,6390%	66,4084%	241	0.06071
80	0,04979	1,24825	2,4896%	62,4125%	241	0.02171
85	0,01660	1,17578	0,8299%	58,7891%	241	0.00642
90	0,00830	1,11091	0,4149%	55,5453%	241	0.00128
95	0,00000	1,05242	0,0000%	52,6212%	241	0.00018
100	0,00000	1,00000	0,0000%	50,0000%	240	0.00003

Tabela 21 – Assessment Score Ranking

7.13) Anexo 13: Gráficos de apoio a Assessment Score Ranking

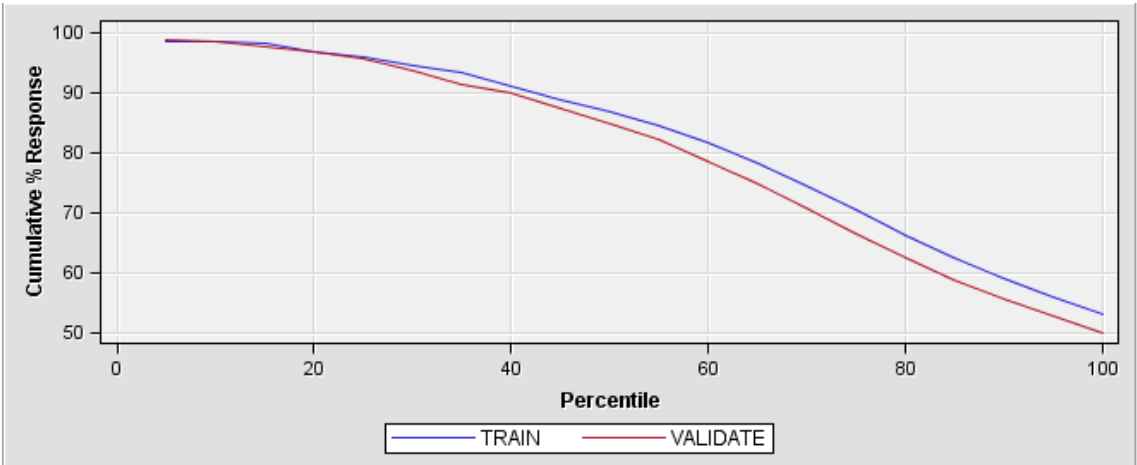


Figura 4 – Gráfico Cumulative % Response do modelo vencedor



Figura 5 – Gráfico de Lift do modelo vencedor

7.14) Anexo 14: Extrapolação de Custo/Benefício a partir de tabela de Assessment Score Ranking

Percentile	Nº de Clientes	Custo da Campanha (5€ p/cliente)	Cumulative % Response	Alvo atingido	Valor Retorno (valor cliente recuperado = 25€ e taxa sucesso campanha = 30%)	Custo/Benefício
5	241	1.204,50 €	98,7552%	238	1.784,26 €	579,76 €
10	482	2.409,00 €	98,5477%	475	3.561,02 €	1.152,02 €
15	723	3.613,50 €	97,6487%	706	5.292,80 €	1.679,30 €
20	964	4.818,00 €	96,7842%	933	6.994,59 €	2.176,59 €
25	1205	6.022,50 €	95,6846%	1153	8.643,91 €	2.621,41 €
30	1445	7.227,00 €	93,7068%	1354	10.158,29 €	2.931,29 €
35	1686	8.431,50 €	91,2270%	1538	11.537,71 €	3.106,21 €
40	1927	9.636,00 €	89,7822%	1730	12.977,12 €	3.341,12 €
45	2168	10.840,50 €	87,4136%	1895	14.214,11 €	3.373,61 €
50	2409	12.045,00 €	84,6409%	2039	15.292,49 €	3.247,49 €
55	2650	13.249,50 €	82,1132%	2176	16.319,38 €	3.069,88 €
60	2891	14.454,00 €	78,4504%	2268	17.008,83 €	2.554,83 €
65	3132	15.658,50 €	74,7765%	2342	17.563,32 €	1.904,82 €
70	3373	16.863,00 €	70,6789%	2384	17.877,87 €	1.014,87 €
75	3614	18.067,50 €	66,4084%	2400	17.997,51 €	-69,99 €
80	3854	19.272,00 €	62,4125%	2406	18.042,21 €	-1.229,79 €
85	4095	20.476,50 €	58,7891%	2408	18.056,93 €	-2.419,57 €
90	4336	21.681,00 €	55,5453%	2409	18.064,16 €	-3.616,84 €
95	4577	22.885,50 €	52,6212%	2409	18.063,94 €	-4.821,56 €
100	4818	24.090,00 €	50,0000%	2409	18.067,50 €	-6.022,50 €

Tabela 22 – Extrapolação de Custo/Benefício a partir de tabela de Assessment Score Ranking

