

Universidade Lusófona de Humanidades e Tecnologias

Licenciatura – Informática de Gestão

Trabalho Fim de Curso

Prof. orientador: Dr. Rui Ribeiro

“Segmentação de clientes no canal Horeca”



Nome	Número
Rui Silva	21001746

Índice

Resumo.....	5
Abstract	5
1. Introdução	7
1.1. Objectivo	7
2. Desenvolvimento.....	7
2.1. Contextualização do Tema	7
2.1.1. O Contexto da indústria Agro-Alimentar e sua importância estratégica.....	7
2.1.2. Indústria das Bebidas Refrescantes Não alcoólicas	7
2.1.3. Estrutura de produção.....	8
2.1.4. Enquadramento do negócio	9
2.1.5. Análise SWOT do sector.....	9
2.2. Perspectiva de Métricas existentes	10
2.2.1. Consumos por habitante	10
2.2.2. Evolução 2008 a 2012 (Europa).....	12
2.2.3. Evolução 2008 a 2012 (Portugal).....	12
2.3. Repartição do mercado das bebidas por segmentos	13
2.4. Consumo em 2012.....	14
2.4.1. Bebidas Não Alcoólicas	14
2.4.2. Tipo de bebidas (com e sem Gás)	14
2.4.3. Sabores	15
2.4.4. Tipo de bebida.....	15
3. Caso Prático	16
3.1. Introdução	16
3.2. Metodologia utilizada.....	17
3.3. Descrição e Análise dos Dados	17
3.3.1. Análise Variáveis – Tipo e Papel	17
3.3.2. Análise Variáveis – Variabilidade.....	18
3.4. Análise Variáveis – Transformação	19
3.4.1. Análise Variáveis – Filtragem.....	19
3.4.2. Análise Variáveis – Valores Omissos	19
3.4.3. Transformação de Variáveis – Tratamento dados	19
3.4.4. Transformação de Variáveis – Aplicação Logaritmo.....	22
3.5. Desenvolvimento dos Clusters	23
3.6. Análise de Perfil dos Segmentos	24
3.7. Oportunidades/campanhas a aplicar sobre os segmentos	26
3.8. Diagrama Final.....	26
4. Conclusões	26

5.	Investigações Futuras	27
6.	Referências Bibliográficas	27
	Anexos.....	28
I -	Implementação de projecto.....	28

Índice de Figuras

Figura 1 - Percentagem de produção nacional	9
Figura 2 - Evolução dos consumos em Portugal	13
Figura 3 - Análise consumo - Bebidas Não Alcoólicas.....	14
Figura 4 - Análise consumo - Bebidas com e sem gás	15
Figura 5 - Análise consumo - Sabores	15
Figura 6 - Análise consumo – Tipo de Bebida.....	16
Figura 7 - StatExplore Inicial	18
Figura 8 - Clusters encontrados/definidos.....	23
Figura 9 - Frequência de registos por segmento.....	23
Figura 10 - Variáveis por segmento e respectiva importância	24
Figura 11 - Importância das variáveis na definição dos Clusters	24

Resumo

Na actual conjuntura económica os mercados de consumo estão a ser afectados, nomeadamente pelas dificuldades que as famílias estão sentir em responder às suas responsabilidades, reflectindo nas vendas e nos respectivos hábitos de consumo. Hábitos que em 2012 ficaram marcados pelo acentuar da falta de confiança do consumidor face ao que já se tinha registado em 2011 e 2010. Os seus comportamentos de compra revelam uma adopção, antecipada e clara, de um ajustamento profundo dos seus consumos como consequência dos rendimentos familiares disponíveis.

Esta situação teve impactos nas empresas em Portugal, ao sofrerem cortes significativos nos volumes de vendas e respectiva facturação não só decorrente da situação económica, mas também do agravamento fiscal. Agravamento este, ainda com mais significado no mercado de bebidas de alta rotação e no sector de restauração, hotéis, cafés e similares (denominado de canal Horeca).

Cada vez mais é previsível que ao longo dos tempos actuais, as empresas tenham quebras nas suas vendas, pelo que é imperioso serem identificados factores relevantes de poupanças e ganhos de produtividade, assim como diferenciação, em relação à concorrência. Neste contexto não é fácil fazer propostas de valor que o cliente e/ou consumidor adopte. Mesmo propostas que seriam reconhecidas como inovadoras e diferenciáveis há anos atrás, hoje têm uma enorme dificuldade em vingar face às atitudes e comportamentos. Situação que tem levado a que o grau de inovação tenha permanecido baixo, pois não existe um retorno adequado ou expectável aos investimentos de inovação.

Por todas estas situações, entre outras aplicadas ao mercado em análise, as empresas tentam fazer análises e estudos comportamentais das necessidades dos mercados, dos clientes e seus hábitos de consumos. Isto para que sejam tomadas decisões de negócio e estratégicas para obter mais valor para a empresa.

Abstract

In nowadays economic situation, the consumption markets are being affected, namely by the difficulties that families are facing in responding to their responsibilities which is affecting the sales and the consumerism habits. These consumerism habits, in 2012, were marked by the accentuating lack of the consumer's confidence when compared with what had been recorded in 2011 and 2010. Their buying habits reveal, clear and in advance, an adoption of a deep adjusting of their consumption as a consequence of the families' available income.

This situation has its impacts in the companies in Portugal as they have gone through significant cuts in their sales and, consequently, in their billing due, not only because of today's economical situation but also because of the tax increase. This increase is even more significant in the beverage market in high rotation/turnover and in the catering, hotels, cafés and similar sectors (called the Horeca's channel).

Nowadays, it is more and more predictable that companies have collapses in their sales and, being so, it is imperative that relevant saving and gains of productivity factors as well as differentiation related to competition may be identified. In this context, it is not easy to make proposals of values which may be

adopted by clients and/or customers. Bearing in mind nowadays' behaviours and attitudes, even proposals which would be seen as innovative and distinguishable years ago, would face serious difficulties in being accepted. Due to this situation, the level of innovation has thus been remained in a very low level as there is not a suitable and expectable re-exchange to the investments.

Due to of all these situations, among other applied to the market analysed here, the companies try to make analyses and behavioural studies of the markets' needs and of the customers and their habits of consumption. All these steps are followed in order to take decisions on business and on strategies so as to obtain more income for the company.

1. Introdução

1.1. Objectivo

O objectivo deste trabalho é, á luz dos conhecimentos adquiridos durante as aulas, o desenvolvimento do modelo comportamental (com foco no modelo de segmentação), com recurso a técnicas de Data Mining, aplicado à indústria de bebidas de alta rotação, para análise de perfil e comportamental, tendências e sazonalidades de clientes no mercado horeca¹. Esta identificação do perfil dos clientes irá permitir conhecer com maior detalhe os comportamentos dos Clientes, de forma a otimizar os seus canais de contacto, permitindo fazer a retenção dos mesmos (através de interacção personalizada), atrair novos clientes e ao mesmo tempo fazer desenvolvimentos para novas oportunidades de negócio.

2. Desenvolvimento

2.1. Contextualização do Tema

2.1.1.O Contexto da indústria Agro-Alimentar e sua importância estratégica

A indústria de bebidas não alcoólicas faz parte da Industria Agro-Alimentar (IAA), sendo um subsector da mesma. Esta indústria, pelo seu impacto directo e indirecto na economia, assume uma elevada importância estratégica no contexto nacional.

IAA – Contributos directos	IAA – Contributos directos	Peso na indústria transformadora
Volume de negócios (M€)	13.400	18%
VAB ² (M€)	2.600	15%
Emprego directo (postos de trabalho)	109.000	15%

Categorias	Volume de negócios (M€)
Bebidas refrigerantes	578,60 ³
Indústria Agro – Alimentar (IAA)	13.400,00 ⁴

2.1.2.Indústria das Bebidas Refrescantes Não alcoólicas

O sector da indústria de bebidas não alcoólicas contempla a produção, comercialização e a distribuição de diversos produtos alimentares, incluindo as bebidas:

- À base de sumos de frutos ou à base de extractos vegetais (colas, chás ou as lima-limão)
- Tónicas
- Energéticas para desportistas

¹ “**Horeca** (or **HORECA**) is the sector of the food service industry that consists of establishments which prepare and serve food and beverages. The term is a syllabic abbreviation of the words Hotel/Restaurant/Catering.” In <http://en.wikipedia.org/wiki/Horeca>

² **Valor acrescentado bruto** (VAB) é o resultado final da actividade produtiva no decurso de um período determinado. Resulta da diferença entre o valor da produção e o valor do consumo intermédio, originando excedentes.

³ Fonte INE - 2009

⁴ Fonte FIPA / Deloitte - 2011

- Bebidas funcionais
- Águas aromatizadas
- Entre outras

Apesar de todas as dificuldades e crise que o país ainda atravessa, com maior incidência em 2012, o sector das bebidas não alcoólicas apresenta uma elevada importância económica, onde o efeito multiplicador desta mesma indústria é reflectido em muitas outras actividades:

- Fornecedores de ingredientes de matérias-primas de fruta
- Fornecedores de material de embalagem
- Fabricantes de equipamentos industriais
- Serviços de manutenção industrial
- Serviços de marketing, de comunicação e media
- Consultorias legais e informáticas
- Serviços de distribuição e comercialização
- Acções e materiais de ponto de venda, etc.

Este sector emprega, como estimativa, 12000 postos de trabalho divididos por:

	Postos de Trabalho ⁵
Emprego directo	3.000
Emprego indirecto	9.000 ⁶

Na estimativa do emprego indirecto, não estão incluídos os impactes sociais ao nível do retalho e da restauração

2.1.3. Estrutura de produção

Em 2011 existiam a nível nacional 23 empresas dedicadas à produção e comercialização de bebidas não alcoólicas. Este número contempla as empresas que produzem bebidas à base de águas de nascente e minerais naturais.

Este sector de bebidas, é reconhecido pela sua presença de empresas nacionais, onde têm no seu portefólio marcas nacionais com forte implantação e notoriedade, coincidindo ou não com marcas internacionais que são produzidas, comercializadas e distribuídas sob licença dos seus proprietários.

Existe ainda, também, um conjunto de empresas que se dedica essencialmente ao fabrico de bebidas refrescantes não alcoólicas para a grande distribuição ou de primeiro preço. As chamadas de “marcas brancas” ou marcas de distribuição.

⁵ Informação disponibilizada pelo Ministério do Emprego, complementada com dados da PROBEB

⁶ Não estão incluídos nesta estimativa os impactes sociais ao nível do retalho e da restauração

A produção nacional tem uma contribuição significativa para o mercado das bebidas refrescantes não alcoólicas. Cerca de 85% da produção de marcas próprias de fabricantes presentes no mercado são produção nacional.

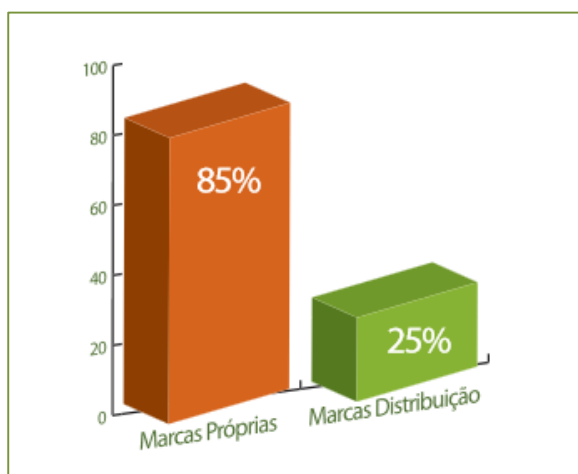


Figura 1 - Percentagem de produção nacional

2.1.4. Enquadramento do negócio

Como já foi referido, o negócio do sector de bebidas de alta rotação está, como maior incidência em 2012, com a maior conjuntura adversa. O desemprego, a política, as estratégias, a incerteza da actual situação económica, a dependência dos mercados externos e o cumprimento das obrigações com as ajudas externas fazem com que este mercado esteja a atravessar dificuldades e isso faz com que as empresas se adaptem a novas realidades e necessidades, sendo o futuro muito incerto.

Esta necessidade de adaptação faz com que existam pontos a considerar:

- Aumentos substanciais dos preços de matérias-primas, embalagens e energia;
- Diminuição do poder de compra em Portugal, o que implicará uma redução do consumo e uma maior relevância do factor preço;
- Mercados internacionais com perspectivas de crescimento.
- Disputa do mercado com outras marcas e empresas poderosas
- Queda do consumo no canal Horeca devido à recessão económica e aumento do IVA

2.1.5. Análise SWOT do sector

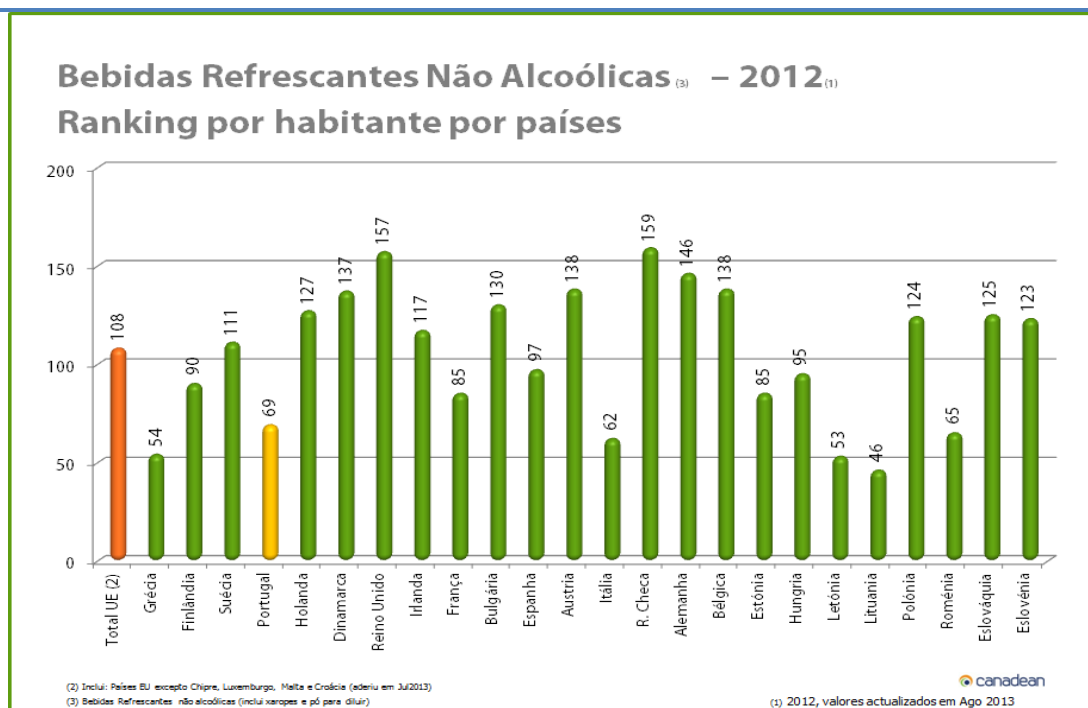
No quadro seguinte podemos verificar alguns pontos que podem fazer a diferença numa empresa que esteja inserida no sector de bebidas.

Análise Interna	Análise Externa
<p><u>Forças</u></p> <ul style="list-style-type: none"> › Marcas com história, especialistas, atestado pela antiguidade e extensão de gama de produtos; › Imagens forte e diferenciada; › Capacidade de Inovação; › Grande cobertura de vendas › Ser um operador de referência no mercado de bebidas não alcoólicas 	<p><u>Oportunidades</u></p> <ul style="list-style-type: none"> › Consumidor encontrará nas marcas factores de diferenciação face às da concorrência; › Clientes reféns de um monopólio encontram-se permeáveis à mudança › Valorização da inovação pelo consumidor › Adequar os produtos aos novos momentos de consumos pela escassez de tempo › Aumento da consciencialização da necessidade de uma boa alimentação
<p><u>Fraquezas</u></p> <ul style="list-style-type: none"> › Desconhecimento de atributos de caracterização dos pontos de venda; › Capacidade investimento para novos lançamentos › Preços de venda 	<p><u>Ameaças</u></p> <ul style="list-style-type: none"> › Enquadramento económico e social › Recessão prolongada e diminuição do consumo; › Elevados custos de produção por falta de matéria-prima › Aumento das taxas de IVA da restauração (por ex.13% para 23%) torna as refeições de lazer mais esporádicas e as diárias mais contidas; › Rápida reprodução da inovação pelas marcas de distribuidor; › Excessiva carga promocional (constante e insustentável) na distribuição moderna por parte dos 2 líderes

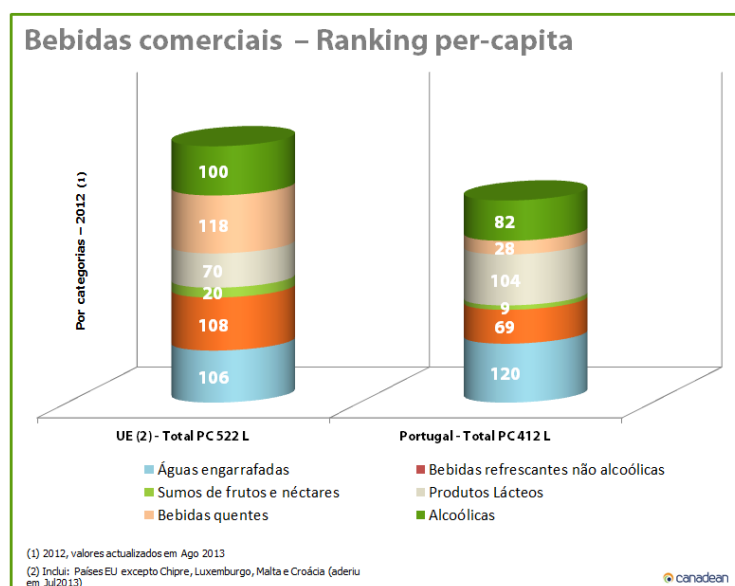
2.2. Perspectiva de Métricas existentes

2.2.1.Consumos por habitante

Em 2012 podemos observar nos quadros em baixo, em Portugal, o consumo por habitante de bebidas refrescantes não alcoólicas é de 69 litros / ano, muito inferior à média da UE de 108 litros / ano.



No âmbito mais alargado do consumo de bebidas comerciais, verificamos um diferencial entre o consumo por habitante de bebidas comerciais em Portugal (412 litros por habitante) por comparação com a média da União Europeia (522 litros / ano), uma diferença de 110 litros por habitante/ano.



No entanto é de salientar que o consumo, em Portugal, de produtos lácteos (104 litros / ano) é mais elevado do que na média da UE (70 litros / ano). Também o consumo de águas engarrafadas em Portugal (120 litros / ano) é mais elevado do que a média da UE (106 litros / ano).

2.2.2. Evolução 2008 a 2012 (Europa)

No quadro seguinte, podemos verificar que, no período entre 2008 e 2012, na Europa Ocidental, o conjunto das bebidas comerciais, teve uma evolução moderada, mas negativa das vendas (-0,6%). Em Portugal podemos observar que apenas em águas engarrafadas, houve uma ligeira evolução (+ 1,4%), e no total a queda foi acentuada, em cerca de - 6,6%.

Bebidas comerciais		
Varição 2008 / 2012 ⁽¹⁾	UE ⁽²⁾	Portugal
Bebidas refrescantes não alcoólicas	+ 2,3 %	- 14,7 %
Águas engarrafadas	+ 0,9 %	+1,4 %
Sumos de frutos e néctares	- 8,1 %	- 8,8 %
Produtos Lácteos <small>(Leite, iogurtes e outros)</small>	- 2,0 %	- 2,5 %
Bebidas quentes <small>(Chá e café)</small>	+ 1,0 %	- 1,6 %
Alcoólicas <small>(Cerveja, vinho e espirituosas)</small>	- 5,5 %	- 15,2 %
Total	- 0,9 %	- 6,6 %

(1) 2012, valores actualizados em Ago 2013
(2) Inclui: Países EU excepto Chipre, Luxemburgo, Malta e Croácia (aderiu em Jul2013)

canadean

Assumindo o tema do trabalho, podemos verificar que a maior diferença de consumo entre 2008 a 2012, entre a EU e Portugal, incide nos 3 maiores decréscimos de consumo: bebidas refrescantes não alcoólicas (- 14,7%), sumos de fruta e néctares (-8,8%) e alcoólicas (-15,2%).

2.2.3. Evolução 2008 a 2012 (Portugal)

Pelo quadro seguinte podemos verificar que a evolução de consumo foi sempre diminuindo ao longo dos anos com as maiores quedas a serem encontradas entre o ano de 2011 e o de 2012. As maiores quedas foram registadas nas bebidas para desportistas (-27,3%), seguida das bebidas sem gás (20,4%), águas engarrafadas (-17,0%) e bebidas de chá e café (-12,9%).

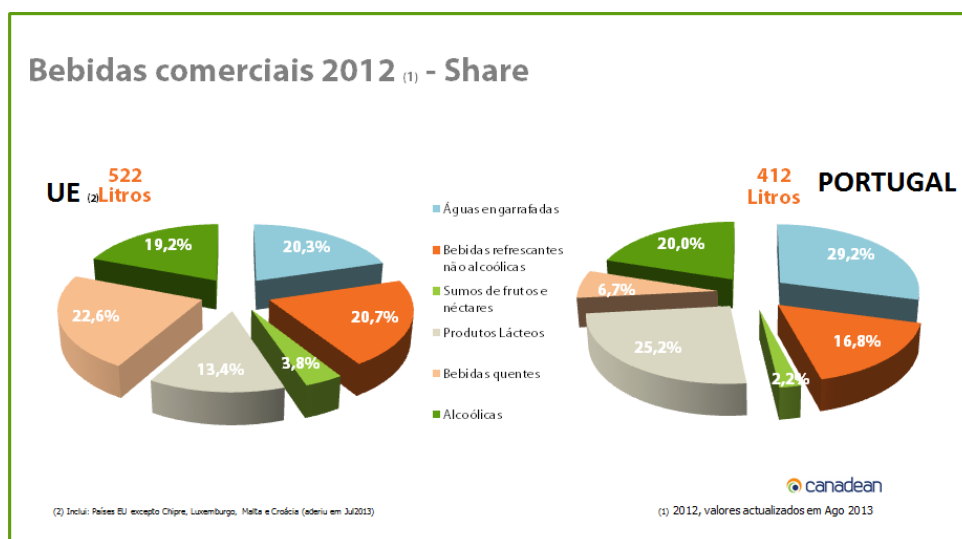
BEBIDAS COMERCIAIS - PORTUGAL	Milhões de Litros					Δ 12/11 em %
	2008	2009	2010	2011	2012 (6)	
TOTAL BEBIDAS REFRESCANTES NÃO ALCÓOLICAS	2.257,3	2.358,3	2.445,1	2.338,1	2.136,7	-8,6
Águas engarrafadas - 10 lt (1)	1.189,4	1.280,1	1.346,6	1.305,6	1.221,7	-6,4
Águas engarrafadas + 10 lt (2)	83,9	83,8	84,8	83,6	69,4	-17,0
Sumos de frutos	27,2	27,6	26,9	24,0	22,8	-5,2
Néctares	84,2	86,1	89,2	81,4	78,8	-3,1
Bebidas concentradas e xaropes (3)	18,1	19,5	20,9	21,9	22,6	3,2
Bebidas à base de pó	23,5	23,1	22,6	22,3	21,7	-2,5
Bebidas refrigerantes gaseificadas	459,8	446,5	442,6	407,6	367,9	-9,8
Bebidas refrigerantes sem gás	170,0	155,8	147,6	125,1	99,6	-20,4
Bebidas de chá café	191,5	226,9	256,2	258,9	225,6	-12,9
Bebidas para desportistas	5,1	4,5	3,4	2,8	2,1	-27,3
Bebidas energéticas	4,6	4,4	4,5	4,8	4,7	-3,7
TOTAL BEBIDAS ALCÓOLICAS	1.045,6	1.009,8	1.000,7	950,8	886,4	-6,8
Cerveja	636,3	612,7	605,9	560,7	505,0	-9,9
Vinho	352,7	342,4	340,9	337,5	330,0	-2,2
Bebidas espirituosas	56,6	54,7	53,9	52,7	51,4	-2,3
TOTAL PRODUTOS LACTEOS	1.147,9	1.143,6	1.148,5	1.134,6	1.118,9	-1,4
Leite branco	932,9	922,6	915,7	902,7	891,7	-1,2
Outros leites (Soja, condensado e fermentado)	94,5	97,9	102,4	104,4	103,7	-0,6
Iogurtes líquidos	120,5	123,2	130,5	127,6	123,5	-3,2
TOTAL BEBIDAS QUENTES	302,8	301,5	300,0	300,3	297,9	-0,8
Café (4)	272,7	271,9	270,3	270,6	268,4	-0,8
Chá (5)	30,1	29,6	29,7	29,7	29,5	-0,7
CONSUMO PER CAPITA BEBIDAS REFRESCANTES NÃO ALCÓOLICAS	211,4	220,2	227,7	217,3	198,2	
CONSUMO PER CAPITA BEBIDAS ALCÓOLICAS	97,9	94,3	93,2	88,4	82,2	
CONSUMO PER CAPITA PRODUTOS LACTEOS	107,5	106,8	106,9	105,4	103,8	
CONSUMO PER CAPITA BEBIDAS QUENTES	28,4	28,2	27,9	27,9	27,6	

Fonte: canadean

Figura 2 - Evolução dos consumos em Portugal

2.3. Repartição do mercado das bebidas por segmentos

Pelo quadro seguinte, verificamos que a liderança em Portugal era no segmento das Águas engarrafadas (29,2%), seguido pelo dos Produtos Lácteos com (25,2%) e pelas bebidas alcoólicas. As bebidas não alcoólicas representavam 16,8%.



2.4. Consumo em 2012

Das diversas análises a seguir indicadas, podemos verificar que o ano de 2012 foi bastante prejudicial para as empresas que estão inseridas no mercado de bebidas.

2.4.1. Bebidas Não Alcoólicas

No período entre 2008 e 2012, houve uma diminuição significativa do consumo de bebidas refrescantes não alcoólicas em Portugal. O consumo reduziu de 844 milhões de litros para 703 milhões de litros, o que representa um decréscimo de 16,7%.



Figura 3 - Análise consumo - Bebidas Não Alcoólicas

* Os valores das bebidas à base de chá/café; isotónicas e energéticas e águas com sabores de 2012 são estimadas

* Nas bebidas refrescantes inclui-se (para além das tradicionais) as seguintes: Tónicas, ginger-ale, soda, bitter, gasosas, isotónicas, energéticas e águas com sabores

2.4.2. Tipo de bebidas (com e sem Gás)

Na análise ao tipo de bebidas (com gás / sem gás) verifica-se um decréscimo em 2012 de 11,7% e 14,3%, respectivamente. Analisando o período dos últimos 5 anos, é de notar que foram as bebidas com gás que mais decresceram, passando de 512 para 386 milhões de litros (-24,6 %).

No caso das bebidas sem gás a queda apesar de acentuada, esta foi menor (-4,8%).

Nas bebidas sem gás o decréscimo foi acentuado entre o ano de 2010 e 2012 (-16,6%).

O aumento do consumo das bebidas sem gás está relacionada, entre outras, com a introdução na legislação da proibição do consumo/venda de bebidas com gás nas escolas por questões de qualidade alimentar e riscos para a saúde.

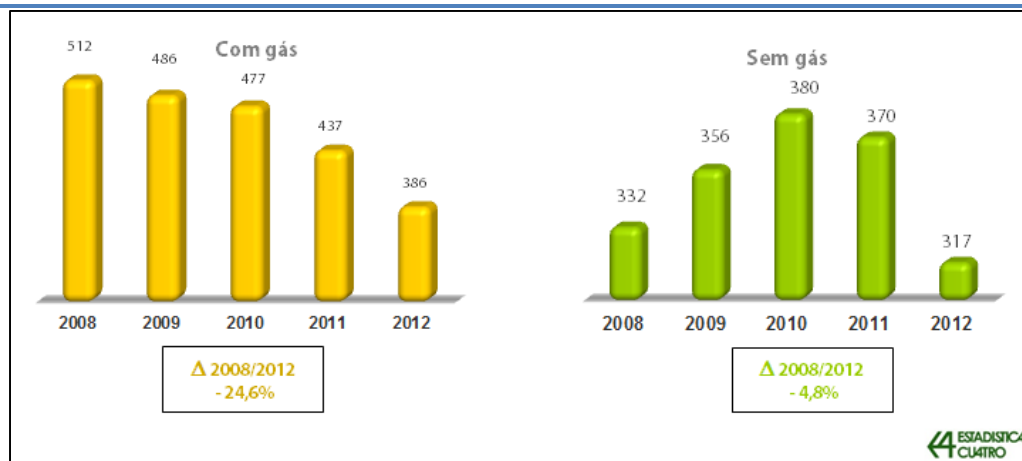


Figura 4 - Análise consumo - Bebidas com e sem gás

2.4.3. Sabores

No consumo por sabores, verifica-se que no mesmo período, houve um decréscimo nos mercados de Colas (-14,7%), de Laranjas (-39,8%) e de Lima-limão (-27,3%). Contrariamente, o mercado de Ice tea / coffee teve um aumento (+21,9%).

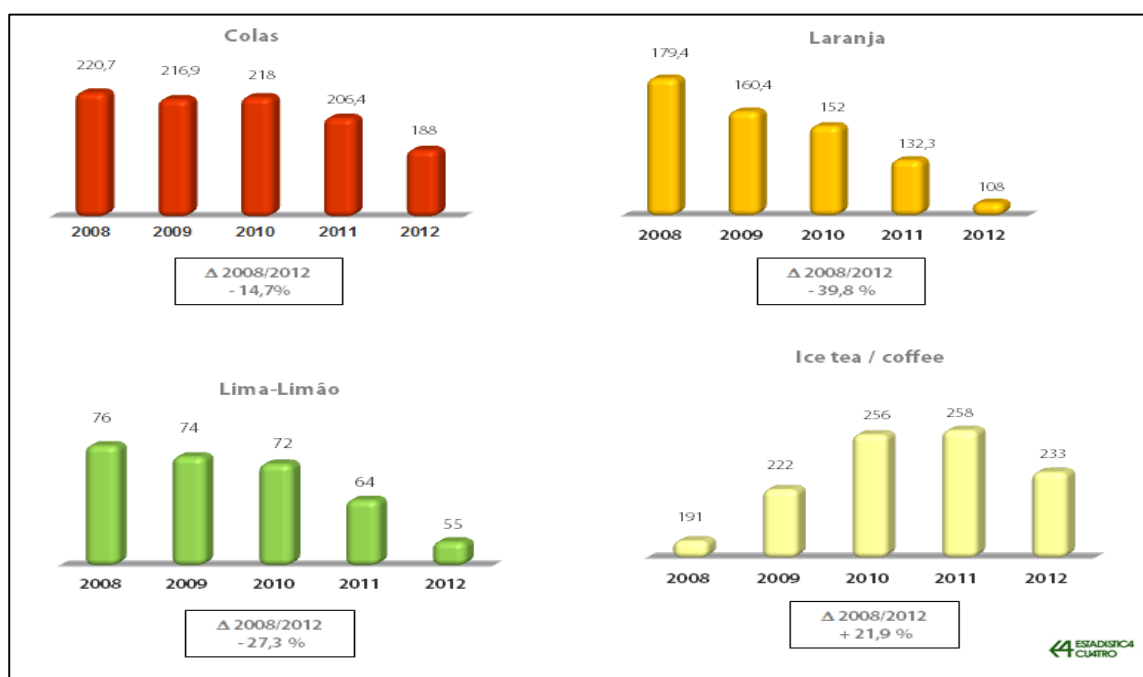


Figura 5 - Análise consumo - Sabores

2.4.4. Tipo de bebida

Fazendo a análise por tipo de bebida, os mercados de bebidas refrescantes não alcoólicas e águas engarrafadas tiveram em decréscimos de 8,6% e 16,7% respectivamente. No caso do mercado de sumos de frutos e néctares, apesar de não ser muito acentuados, houve um decréscimo de 8,6%

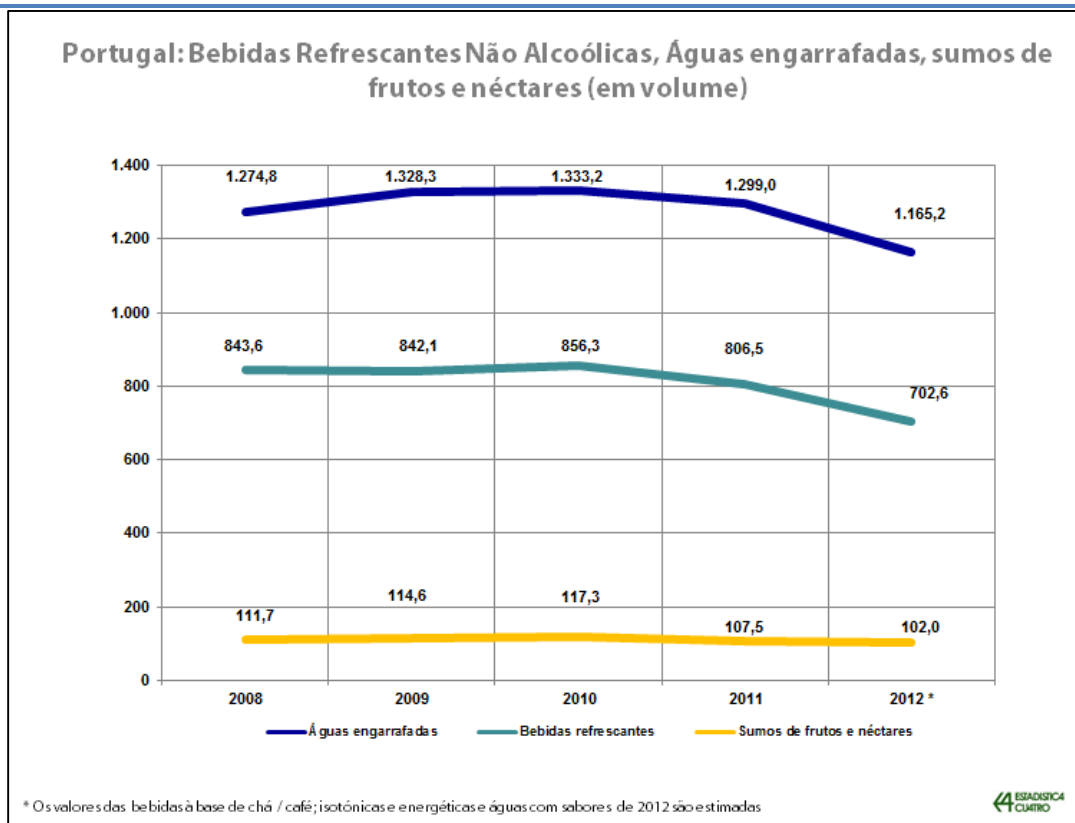


Figura 6 - Análise consumo – Tipo de Bebida

3. Caso Prático

3.1. Introdução

A selecção de dados foi feita em consumos efectuados pelos clientes, em volume, no canal de distribuição Horeca, num período de 12 meses, compreendido entre Julho de 2012 e Junho de 2013, com base em cada um dos modelos de negócio da empresa. Modelos de negócio esses divididos em:

- Nutrição (bebidas de Sumo, Sumos e Néctares, doses fruta)
- Refrigerantes
- Águas
- Cervejas

Para complemento da informação foram adicionados valores totais de consumo, em volume de algumas marcas que fazem parte dos modelos de negócio atrás indicados.

Pela dificuldade de validação de dados, com a caracterização dos pontos de venda, a análise incide em aproximadamente 1500 pontos de venda, de um universo de 30000, existentes na empresa inseridos neste canal de distribuição.

Até ao ser construído o ficheiro final em excel, foram importados dados, mestre de clientes, de volume total de vendas por marca e área de negócio, e de atributos dos pontos de venda, para Microsoft Access. Aqui foram criados reports /query's directos à base dados, sendo depois os resultados exportados para Microsoft Excel já com a informação já normalizada.

3.2. Metodologia utilizada

A metodologia utilizada foi a criação de um modelo de segmentação, com recurso a técnicas de *Data Mining*, aplicado aos dados de consumo de pontos de venda de uma empresa, de forma a serem identificados os principais segmentos de perfis e assim perceber os seus comportamentos actuais, para que regularmente a Empresa esteja preparada para otimizar e ajustar os seus canais de contacto, para cada um dos perfis.

A construção desse modelo, para a análise de resultados e respectivas acções sobre os mesmos, foi feita com a aplicação SAS Enterprise Miner Workstation 12.1.

Para isso houve a necessidade de criar uma Analytical Base Table, normalizada para o SAS Miner, com o software JMP Pro 10, após a importação de dados de um ficheiro base de Excel.

3.3. Descrição e Análise dos Dados

A ABT⁷ em análise é composta por:

. N. de observações/registos: 1555

. N. de variáveis: 42

3.3.1. Análise Variáveis – Tipo e Papel

Após a importação inicial da ABT e a criação do respectivo diagrama foi necessário proceder à análise de variáveis. O objectivo foi identificar e seleccionar o tipo de papel que cada variável representa no modelo desenvolvido. Assim a ABT, denominada de desenvolvimento, é definida por:

Name	Description	Role	Level
Cliente	Cod. De Cliente	ID	Nominal
Canal	Canal de distribuição	Input	Nominal
Validado	Status validação	Input	Ordinal
Tipologia	Tipologia de cliente	Input	Ordinal
Regiao	Região de cliente	Input	Ordinal
Zona	Zona do Cliente	Input	Nominal
Grupo	Grupo de clientes	Input	Ordinal
AreaCli	Área destinada aos clientes	Input	Interval
CompBalc	Comprimento de balcão	Input	Interval
NLugares	Número de lugares	Input	Interval
NMinCli	Número mínimo de clientes	Input	Interval
NMColabor	Número médio colaboradores	Input	Interval
NMaxCli	Número máximo de clientes	Input	Interval
RealSumos13	Volume total Sumos 13	Input	Interval
RealRefrig13	Volume total Refrigerantes 13	Input	Interval
RealAguas13	Volume total Águas 13	Input	Interval
RealCervejas13	Volume total Cervejas 13	Input	Interval
SumOfDNN12	Volume total Área DNN nos últimos 12 meses	Input	Interval

⁷ Analytical Base Table

SumOfDNR12	Volume total Área DNR nos últimos 12 meses	Input	Interval
SumOfDNAC12	Volume total Área DNAC nos últimos 12 meses	Input	Interval
SumOfDNMS12	Volume total Área DNMS nos últimos 12 meses	Input	Interval
SumOfCP12	Volume total de CP nos últimos 12 meses	Input	Interval
SumOfCE12	Volume total de CE nos últimos 12 meses	Input	Interval
SumOfSU12	Volume total de SU nos últimos 12 meses	Input	Interval
SumOfPP12	Volume total de PP nos últimos 12 meses	Input	Interval
SumOfGA12	Volume total de GA nos últimos 12 meses	Input	Interval
SumOfGT12	Volume total de GT nos últimos 12 meses	Input	Interval
SumOfDC12	Volume total de DC nos últimos 12 meses	Input	Interval
SumOfCT12	Volume total de CT nos últimos 12 meses	Input	Interval
SumOfTG12	Volume total de TG nos últimos 12 meses	Input	Interval
SumOfGR12	Volume total de GR nos últimos 12 meses	Input	Interval
SumOfSE12	Volume total de SE nos últimos 12 meses	Input	Interval
SumOfBE12	Volume total de BE nos últimos 12 meses	Input	Interval
SumOfUB12	Volume total de UB nos últimos 12 meses	Input	Interval
SumOfSM12	Volume total de SM nos últimos 12 meses	Input	Interval
SumOfFR12	Volume total de FR nos últimos 12 meses	Input	Interval
SumOfUC12	Volume total de UC nos últimos 12 meses	Input	Interval
SumOfGL12	Volume total de GL nos últimos 12 meses	Input	Interval
SumOfCL12	Volume total de CL nos últimos 12 meses	Input	Interval
SumOfSN12	Volume total de SN nos últimos 12 meses	Input	Interval
SumOfDM12	Volume total de DM nos últimos 12 meses	Input	Interval
SumOfLT12	Volume total de LT nos últimos 12 meses	Input	Interval

3.3.2. Análise Variáveis – Variabilidade

Após uma análise inicial de variáveis, com o nó de “STAT EXPLORE”, houve a necessidade com base na variabilidade de valores, de alterar o *level* da variável Zona de “Nominal” para Interval”

37	Data Role=TRAIN								
38									
39				Number					
40	Data	Variable		of					
41	Role	Name	Role	Levels	Missing	Mode	Mode Percentage	Mode2	Mode2 Percentage
42									
43	TRAIN	Canal	INPUT	1	0	10	100.0		0.00
44	TRAIN	Regiao	INPUT	1	0	2	100.0		0.00
45	TRAIN	Tipologia	INPUT	9	0	2	32.41	1	28.94
46	TRAIN	Validado	INPUT	3	0	5	52.73	N	37.49
47	TRAIN	Zona	INPUT	18	0	AV	47.97	AU	39.10
48									
49									
50									

Figura 7 - StatExplore Inicial

A análise da variabilidade das variáveis foi feita com as seguintes regras:



3.4. Análise Variáveis – Transformação

3.4.1. Análise Variáveis – Filtragem

Na análise efectuada não foram identificadas situações de necessidade de fazer filtragem de dados.

3.4.2. Análise Variáveis – Valores Omissos

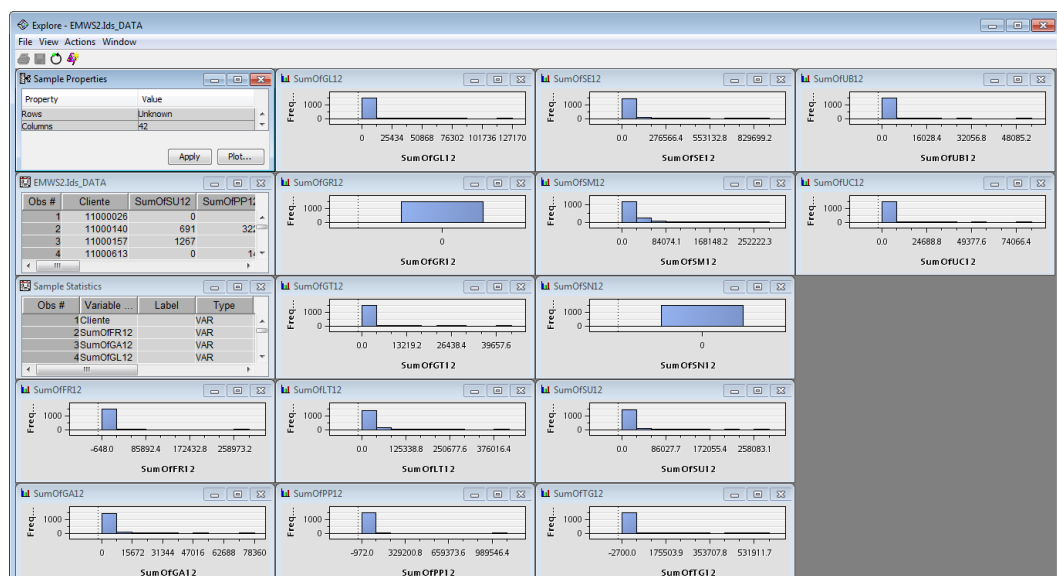
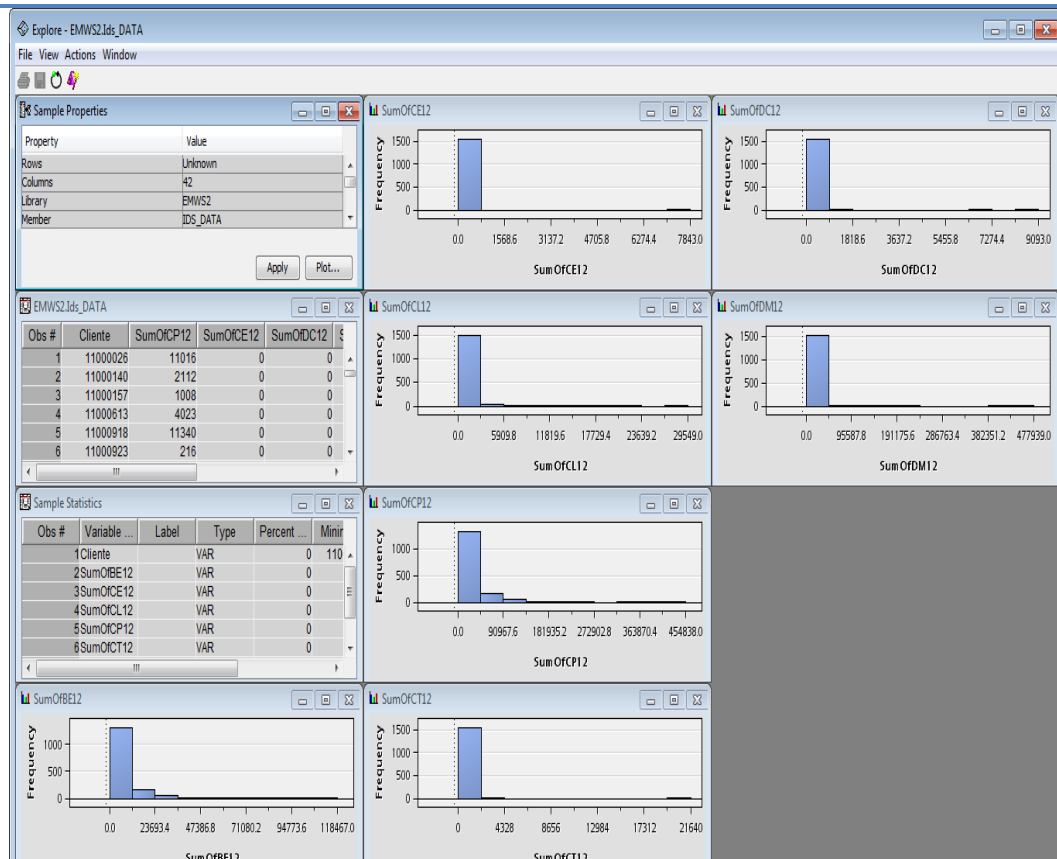
Nos dados recolhidos e importados para o modelo não foram detetados, em qualquer variável, valores omissos. Por este facto não houve a necessidade de fazer qualquer tipo de tratamento. As imagens seguintes demonstram a inexistência de valores omissos.

39									
40	Data	Variable		Number					
41	Role	Name	Role	of	Missing	Mode	Mode	Mode2	Mode2
42				Levels			Percentage		Percentage
43	TRAIN	Canal	INPUT	1	0	10	100.0		0.00
44	TRAIN	Regiao	INPUT	1	0	2	100.0		0.00
45	TRAIN	Tipologia	INPUT	9	0	2	32.41	1	28.94
46	TRAIN	Validado	INPUT	3	0	S	52.73	N	37.49
47	TRAIN	Zona	INPUT	18	0	AV	47.97	AU	39.10
48									
49									

Results - Node: StatExplore.Inicial Diagram: Horeca						
File Edit View Window						
Output						
55	Variable	Role	Mean	Standard	Non	Missing
56				Deviation	Missing	
57	AreaCli	INPUT	48.20257	28.44402	1555	0
58	CompBalc	INPUT	4.688746	4.03921	1555	0
59	Grupo	INPUT	101193	109.6999	1555	0
60	NLugares	INPUT	40.67138	28.97382	1555	0
61	NMColabor	INPUT	3.77492	2.740753	1555	0
62	NMaxCli	INPUT	264.2534	178.4329	1555	0
63	NMinCli	INPUT	126.4752	103.4122	1555	0
64	RealAgua13	INPUT	336.29	864.997	1555	0
65	RealCerveja13	INPUT	46.40772	372.4962	1555	0
66	RealRefrig13	INPUT	379.89	613.2384	1555	0
67	RealSumos13	INPUT	168.7666	335.0813	1555	0
68	SumOfEE12	INPUT	6215.706	10030.06	1555	0
69	SumOfCE12	INPUT	5.04373	198.8919	1555	0
70	SumOfCL12	INPUT	457.0707	1688.406	1555	0
71	SumOfCP12	INPUT	24144.23	34525.36	1555	0
72	SumOfCT12	INPUT	16.05916	553.1945	1555	0
73	SumOfDC12	INPUT	11.72412	296.7434	1555	0
74	SumOfDM12	INPUT	3010.984	22716.79	1555	0
75	SumOfDNAC12	INPUT	47561.85	105772.3	1555	0
76	SumOfDNMS12	INPUT	3339.03	9884.367	1555	0
77	SumOfDNM12	INPUT	24532.57	35267.99	1555	0
78	SumOfDNR12	INPUT	64200.92	85573.52	1555	0
79	SumOfFR12	INPUT	3063.104	8731.873	1555	0
80	SumOfGA12	INPUT	2246.005	4817.816	1555	0
81	SumOfGL12	INPUT	847.8707	6488.148	1555	0
82	SumOfGR12	INPUT	0	0	1555	0
83	SumOfGT12	INPUT	97.36592	1566.758	1555	0
84	SumOfLT12	INPUT	18100.6	27604.82	1555	0
85	SumOfFP12	INPUT	5935.734	33341.88	1555	0
86	SumOfSE12	INPUT	38054.73	81267.49	1555	0
87	SumOfSH12	INPUT	21694.9	26604.18	1555	0
88	SumOfSU12	INPUT	0	0	1555	0
89	SumOfTG12	INPUT	9882.808	17904.22	1555	0
90						
91						
92	SumOfTG12	INPUT	3433.02	30630.68	1555	0
93	SumOfUB12	INPUT	357.0559	2155.233	1555	0
94	SumOfUC12	INPUT	1517.251	5647.527	1555	0
95						

3.4.3. Transformação de Variáveis – Tratamento dados

As variáveis de volume total, denominadas “SumOf...”, com excepção das “SumOfD...” apresentam um elevado nº observações com o valor 0. Assim, foi considerado importante a transformação dessas variáveis de forma a serem mais relevantes para a análise. Na seguinte imagem é possível verificar o estado inicial das variáveis que justificaram a sua transformação:



Assim numa primeira fase, com a utilização do node “SAS Code”, os valores das variáveis em questão foram tratados dando origem a novas variáveis.

Numa segunda fase foi utilizado o node “Metadata”, para converter as novas variáveis criadas, alterando o tipo de dados de intervalar para binário.

Em simultâneo converteram-se as variáveis para intervalares, pois tinham sido definidas como ordinais de forma a permitir a análise e criação das variáveis binárias. Esta transformação permite manter inalterados os valores para uma análise futura.

A lógica utilizada (SAS CODE) é representado pelo código:

```
data &EM_EXPORT_TRAIN;
set &EM_IMPORT_DATA;
if (SumOfBE12 <= 0) then tem_SumOfBE12 = 0;
else tem_SumOfBE12 = 1;

if (SumOfCE12 <= 0) then tem_SumOfCE12 = 0;
else tem_SumOfCE12 = 1;

if (SumOfCL12 <= 0) then tem_SumOfCL12 = 0;
else tem_SumOfCL12 = 1;

if (SumOfCP12 <= 0) then tem_SumOfCP12 = 0;
else tem_SumOfCP12 = 1;

if (SumOfCT12 <= 0) then tem_SumOfCT12 = 0;
else tem_SumOfCT12 = 1;

if (SumOfDC12 <= 0) then tem_SumOfDC12 = 0;
else tem_SumOfDC12 = 1;

if (SumOfDM12 <= 0) then tem_SumOfDM12 = 0;
else tem_SumOfDM12 = 1;

if (SumOfFR12 <= 0) then tem_SumOfFR12 = 0;
else tem_SumOfFR12 = 1;

if (SumOfGA12 <= 0) then tem_SumOfGA12 = 0;
else tem_SumOfGA12 = 1;

if (SumOfGL12 <= 0) then tem_SumOfGL12 = 0;
else tem_SumOfGL12 = 1;

if (SumOfGR12 <= 0) then tem_SumOfGR12 = 0;
else tem_SumOfGR12 = 1;

if (SumOfGT12 <= 0) then tem_SumOfGT12 = 0;
else tem_SumOfGT12 = 1;

if (SumOfLT12 <= 0) then tem_SumOfLT12 = 0;
else tem_SumOfLT12 = 1;

if (SumOfPP12 <= 0) then tem_SumOfPP12 = 0;
else tem_SumOfPP12 = 1;

if (SumOfSE12 <= 0) then tem_SumOfSE12 = 0;
else tem_SumOfSE12 = 1;

if (SumOfSM12 <= 0) then tem_SumOfSM12 = 0;
else tem_SumOfSM12 = 1;

if (SumOfSN12 <= 0) then tem_SumOfSN12 = 0;
else tem_SumOfSN12 = 1;

if (SumOfSU12 <= 0) then tem_SumOfSU12 = 0;
else tem_SumOfSU12 = 1;

if (SumOfTG12 <= 0) then tem_SumOfTG12 = 0;
else tem_SumOfTG12 = 1;

if (SumOfUB12 <= 0) then tem_SumOfUB12 = 0;
else tem_SumOfUB12 = 1;

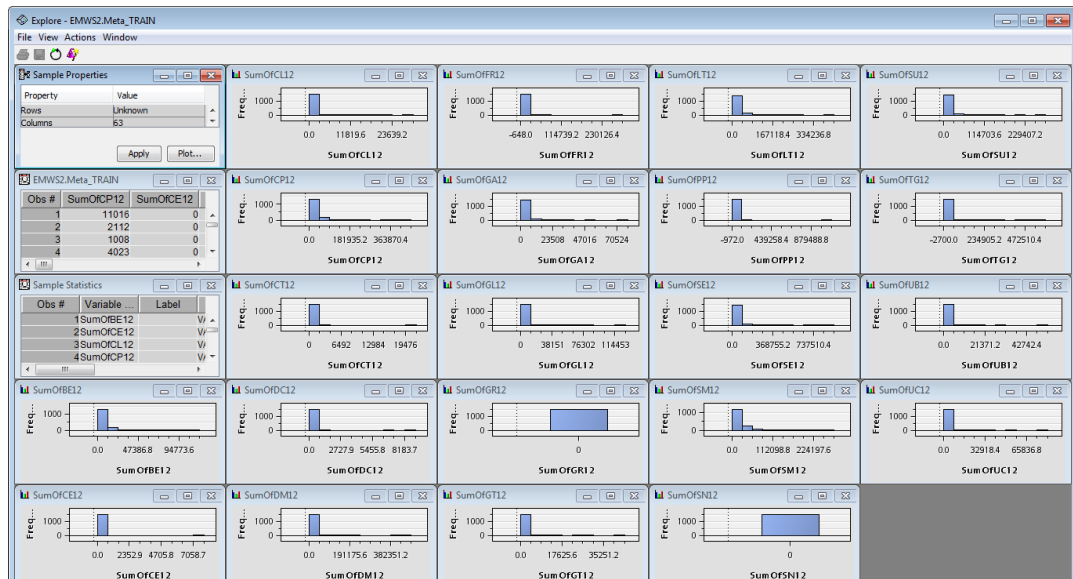
if (SumOfUC12 <= 0) then tem_SumOfUC12 = 0;
else tem_SumOfUC12 = 1;

run;
```

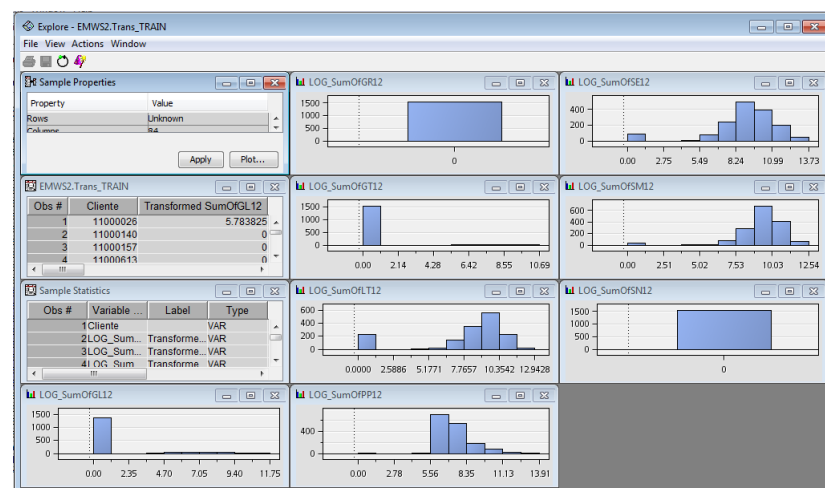
3.4.4. Transformação de Variáveis – Aplicação Logaritmo

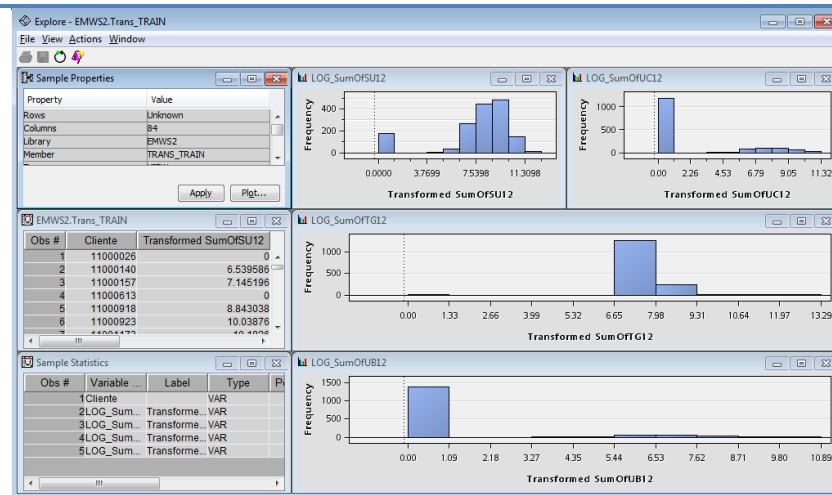
Depois de uma análise dos intervalos de volumes, em termos de assimetria, foi tomada a decisão de aplicar um logaritmo de forma a suavizar as assimetrias.

Estado inicial dos intervalos das variáveis dos volumes:



Estados das variáveis dos intervalos de volumes, em termos de assimetria, após o tratamento:





Esta transformação permite que a distribuição de dados seja mais heterogénea e que os segmentos sejam mais significativos para análise.

3.5. Desenvolvimento dos Clusters

Na aplicação do node de “Cluster”, com os valores por omissão, verificou-se que eram gerados 19 clusters.

Nas opções do node “Cluster” o tipo de standardização interna foi alterado para o modo *range*, sendo este o modo indicado quando existem atributos quantificados em escalas diferentes.

Depois de diversas tentativas, para obter o número óptimo de clusters, fo encontrado o valor de 5 clusters, com é possível verificar na imagem seguinte:

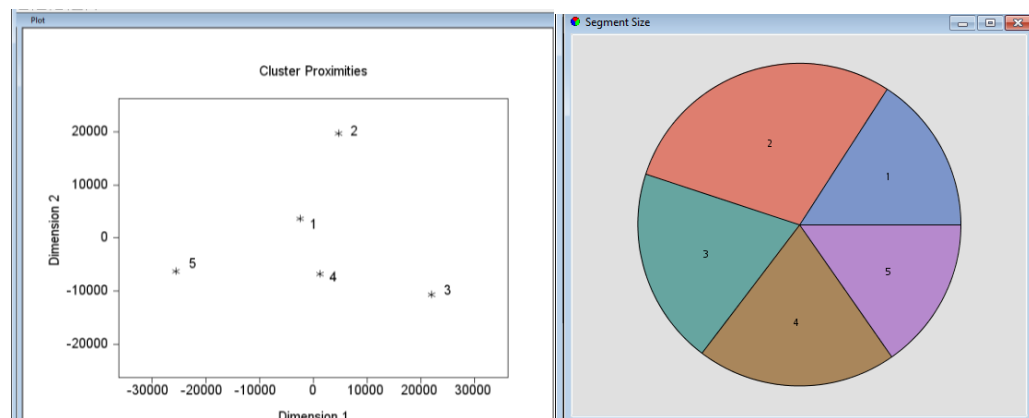


Figura 8 - Clusters encontrados/definidos

Frequencies: _SEGMENT_			
Segment Variable	Segment Value	Frequency Count	Percent of Total Frequency
SEGMENT	2	452	29.0675
SEGMENT	4	310	19.9357
SEGMENT	3	306	19.6785
SEGMENT	1	248	15.9486
SEGMENT	5	239	15.3698

Figura 9 - Frequência de registos por segmento

Variable: _SEGMENT_Segment: 1 Count: 248 Decision Tree Importance Profiles			Variable: _SEGMENT_Segment: 2 Count: 452 Decision Tree Importance Profiles			Variable: _SEGMENT_Segment: 3 Count: 306 Decision Tree Importance Profiles		
Variable	Worth	Rank	Variable	Worth	Rank	Variable	Worth	Rank
Tipologia	0.25565	1	Tipologia	0.41012	1	Tipologia	0.31612	1
NMColabor	0.01810	2	LOG_SumOfCP12	0.06607	2	SumOfDNN12	0.05344	2
Zona	0.01340	3	SumOfDNN12	0.06421	3	LOG_SumOfCP12	0.05213	3
NLugares	0.01276	4	AreaCli	0.04416	4	NMinCli	0.02535	4
AreaCli	0.01196	5	LOG_SumOfSU12	0.04302	5	NLugares	0.01747	5
LOG_SumOfSU12	0.01066	6	NLugares	0.02864	6	NMColabor	0.01427	6
LOG_SumOfUC12	0.00896	7	NMColabor	0.02792	7	Zona	0.00617	7
LOG_SumOfCP12	0.00602	8	NMinCli	0.02150	8	LOG_SumOfUC12	0.00576	8
SumOfDNN12	0.00549	9	LOG_SumOfUC12	0.01429	9	LOG_SumOfSU12	0.00513	9
NMinCli	0.00372	10	Zona	0.00870	10	AreaCli	0.00354	10

Variable: _SEGMENT_Segment: 4 Count: 310 Decision Tree Importance Profiles			Variable: _SEGMENT_Segment: 5 Count: 239 Decision Tree Importance Profiles		
Variable	Worth	Rank	Variable	Worth	Rank
Tipologia	0.14137	1	Tipologia	0.086478	1
Zona	0.08285	2	Zona	0.051236	2
SumOfDNN12	0.00941	3	NLugares	0.016225	3
LOG_SumOfCP12	0.00778	4	AreaCli	0.016001	4
NMinCli	0.00768	5	NMColabor	0.012529	5
NMColabor	0.00536	6	LOG_SumOfSU12	0.007683	6
LOG_SumOfSU12	0.00511	7	LOG_SumOfCP12	0.006618	7
LOG_SumOfUC12	0.00493	8	SumOfDNN12	0.006483	8
NLugares	0.00394	9	NMinCli	0.005503	9
AreaCli	0.00266	10	LOG_SumOfUC12	0.003993	10

Figura 10 - Variáveis por segmento e respectiva importância

3.6. Análise de Perfil dos Segmentos

Como procedimento é aconselhável que as variáveis que mais se destacam, ou as mais relevantes, sejam retiradas da análise. Neste caso a variável “Tipologia”, com uma importância de 100% no Cluster e respectiva importância destacada em todos os segmentos deveria ser retirada. Aqui esta variável foi mantida por opção pois assume valores diferentes em cada segmento e a tipologia de um cliente está relacionada com o consumo de bebidas por divisão de negócio.

Para a extracção de perfis para a segmentação foram consideradas as variáveis com valores superiores a 50% sendo rejeitadas as restantes com valores inferiores.

NAME	LABEL	NRULES	NSURROGATES	IMPORTANCE
Tipologia		4	0	1.00000
LOG_SumOfCP12	Transformed SumOfCP12	0	2	0.74548
SumOfDNN12		0	2	0.74511
NMColabor		0	2	0.67066
AreaCli		0	3	0.66749
NMinCli		0	2	0.65549
Zona		4	0	0.54148
LOG_SumOfSU12	Transformed SumOfSU12	0	1	0.52953
LOG_SumOfUC12	Transformed SumOfUC12	0	2	0.51959
NLugares		0	1	0.51956
Grupo		0	2	0.42269
LOG_SumOfFP12	Transformed SumOfFP12	0	2	0.42176
SumOfDNN12		0	2	0.42135
NMaxCli		0	2	0.41847
LOG_SumOfSE12	Transformed SumOfSE12	0	1	0.41335
SumOfDNN12		0	1	0.41259
RealRefrigt13		0	1	0.39238
LOG_SumOfIT12	Transformed SumOfIT12	0	1	0.08564
LOG_SumOfGAl2	Transformed SumOfGAl2	0	1	0.08527
LOG_SumOfBE12	Transformed SumOfBE12	0	1	0.05577
RealSumos13		0	1	0.05299

Figura 11 - Importância das variáveis na definição dos Clusters

Após a aplicação do node “Segment Profile” podemos verificar que existem 5 segmentos distintos, conforme mostram as seguintes imagens.



Da análise dos segmentos foram retiradas as seguintes conclusões de perfis:

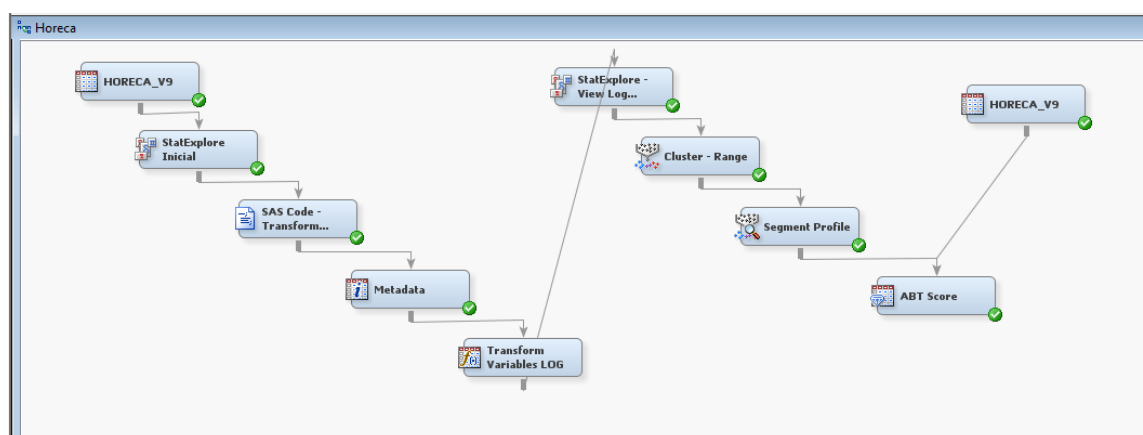
Segmento	Descrição
1	Clientes da tipologia café em zona “Residencial” e “Comércio e Serviços”, com n.º de colaboradores, n.º de lugares, área de clientes e volume de consumo do produto SU superior à média
2	Clientes da tipologia “Restaurante” com área de clientes e n.º de lugares superiores à média. O volume de consumo do produto CP, SU e de produtos de consumo da área de negócio DNN superior à média.
3	Clientes de tipologia “Pastelaria/confeitaria”, com volume de consumo da área de negócio DNN e volume de consumo de CP ambos inferiores à média. Tem n.º de colaboradores, lugares e clientes também inferior à média.
4	Clientes de tipologia “Snack bar” com zona em “Comércio e Serviços” e “Centro Comercial”
5	Clientes de tipologia “Snack bar” com zona “Residencial”. Tem n.º de lugares inferior à média e Área de cliente e n.º colaboradores superior à média.

3.7. Oportunidades/campanhas a aplicar sobre os segmentos

Segmento	Descrição
1	Fomentar o consumo de bebidas com a criação de menus para rentabilização de recursos existentes
2	Com uma forte implementação e de consumo de bebidas da área DNN, devem ser desenvolvidos todos os esforços para uma aposta no consumo de outras unidades de negócio.
3	Sendo um segmento de tipologia “pastelaria/confeitaria”, devem ser desenvolvidos todos os esforços para um maior consumo de bebidas de DNN. Tentar rentabilizar espaços com esplanadas.
4	Oportunidade de divulgação de novos produtos devido a alta rotação e consumo de bebidas em refeições rápidas.
5	Como os locais têm n.º de lugares reduzidos mas a área de cliente ser superior, avaliar a hipótese de colocação de material publicitário

3.8. Diagrama Final

A imagem seguinte mostra o diagrama final, resultado do processo de análise.



4. Conclusões

Como interpretação dos resultados obtidos, a variável Tipologia é a que tem maior importância na definição dos segmentos encontrados para os dados disponíveis. O consumo do produto CP e de produtos que pertencem à Área de Negócio DNN, são também relevantes para a construção do modelo de segmentação e aplicação das respectivas ações para criar novas oportunidades de negócio, manter e/ou aumentar as já existentes.

Os resultados dos atributos que atualmente são utilizados para definir um ponto de venda são muito heterogêneos nos clusters, o que significa a necessidade de realização de mais segmentações com maior granularidade de dados.

Sobre o trabalho houve dificuldades inerentes ao desenvolvimento de um modelo e/ou processo. Neste, em concreto, houve a dificuldade acrescida da obtenção de dados e da dificuldade de utilização da ferramenta SAS Miner que além da complexidade da mesma, o tempo de intervalo em que houve uma aprendizagem e a utilização da mesma para o desenvolvimento do trabalho foi bastante elevado. Houve por isso a necessidade a recorrer a documentação das aulas, apontamentos e trabalhos, assim como diversas pesquisas no site do SAS e respectiva documentação.

5. Investigações Futuras

Como possibilidades futuras temos:

- Cálculo do potencial dos clientes para que um ponto de venda seja segmentado com o seu potencial versus o real;
- Desenvolvimento de segmentos com a utilização de dados dos consumos do(s) último(s) meses (mensal, último trimestre, último semestre, ano móvel)
- Desenvolver acções de marketing futuras mais eficientes para a retenção e fidelização de clientes.

6. Referências Bibliográficas

- [1] SAS Institute Inc., SAS/STAT® 9.2 User's Guide The SCORE Procedure (Book Excerpt), SAS Institute Inc., 2012
- [2] <http://support.sas.com/resources/thirdpartysupport/v92m3/appservers/jbossw.html>
- [3] <http://support.sas.com/documentation/installcenter/en/ikcampaignmgr/64366/PDF/default/sreq.pdf>
- [4] <http://support.sas.com/resources/thirdpartysupport/v92m3/appservers/webspheresw.html>
- [5] <http://support.sas.com/resources/thirdpartysupport/v92m3/appservers/weblogicsw.html>
- [6] http://www.sas.com/offices/europe/portugal/formacao/calendario_formacao.html
- [7] Ana Margarida Alexandre, Data Mining Aula 8 - Modelos de Segmentação, Informática de Gestão 2012/2013.
- [8] Ana Margarida Alexandre, Data Mining Aula 10 - Modelo de Segmentação – Case Study Retalho Informática e Gestão, Outubro 2012.
- [9] Sas Enterprise Miner 7.1, Reference Help, 2011
- [10] José Verissimo, Luis Ferreira, Rui Silva, Gestão de Projecto - Segmentação de clientes no canal Horeca, Junho de 2013
-

Anexos

I - Implementação de projecto

1. Restrições

De modo a que o projecto, durante as suas várias fases, cumpra com todas as expectativas, prazos e orçamentos, é importante que sejam consideradas, nesta fase, as restrições de diferentes níveis que descrevemos e caracterizamos nos pontos seguintes.

1.1 Organização

Tendo em conta a conjuntura e em particular o mercado em que a organização se insere, uma das restrições mais relevantes são razões económicas e financeiras. Assim, tendo em conta este aspecto e não impedindo o cliente de adquirir o sistema pretendido com a qualidade desejada, o preço apresentado para a solução final é uma oportunidade que está integrada na nossa oferta promocional internacional para este semestre e só é válida para esta proposta. Assim, a janela de implementação terá de ser a definida para que se possam assumir e manter os valores em causa.

1.2 Sociais

Tendo em conta que se trata de um sistema informático novo para a generalidade dos utilizadores da organização e principalmente para os utilizadores-alvo que irão utilizar de forma mais intensiva, existirá um esforço por parte da equipa de projecto no sentido de minimizar os riscos da não-aceitação da solução, permitindo criar um modelo simples, usável e intuitivo.

1.3 Tecnológicas

A principal restrição tecnológica inerente a este projecto relaciona-se directamente com o facto de o sistema ser integrado com outros sistemas já existentes na organização. A integração é feita, naturalmente com recurso a API's (Application Programming Interface) fornecidas por esses sistemas.

1.4 Pressupostos / Resultados

É expectável que com a implementação de este projecto, seja possível obter a identificação dos segmentos por tipo de cliente, segmentos esses que se caracterizam por ser heterogéneos entre grupos e homogéneos dentro do mesmo grupo. A partir da identificação e análise desses segmentos será possível:

- Numa primeira fase:
 - Maximizar o conhecimento do perfil e necessidades dos clientes o avaliar tendências por tipologia
 - Criar possíveis oportunidades de negócio
- Futuramente (como trabalho futuro):

-
- Desenvolver acções de marketing futuras mais eficientes para a retenção e fidelização de clientes.

1.5 Orçamento

Valor disponível e orçamentado, aprovado em plano anual, para o projecto – até 200.000,00€.

2. Solução tecnológica

2.1 Módulos SAS

Tendo em conta a estratégia da empresa e a aposta do desenvolvimento de todas as suas soluções com base na mesma plataforma, assume vantagens muito consideráveis para os seus clientes que se materializam em aspectos tão relevantes com processos de manutenção simples, curva de aprendizagem mais rápida, melhor rentabilização de recursos e mais fácil disponibilidade de técnicos com competências adequadas.

Assim, a solução *Customer Analytics* tem no seu conteúdo três módulos, sendo que cada um deles tem características e funcionalidades distintas e funcionam de forma totalmente integrada:

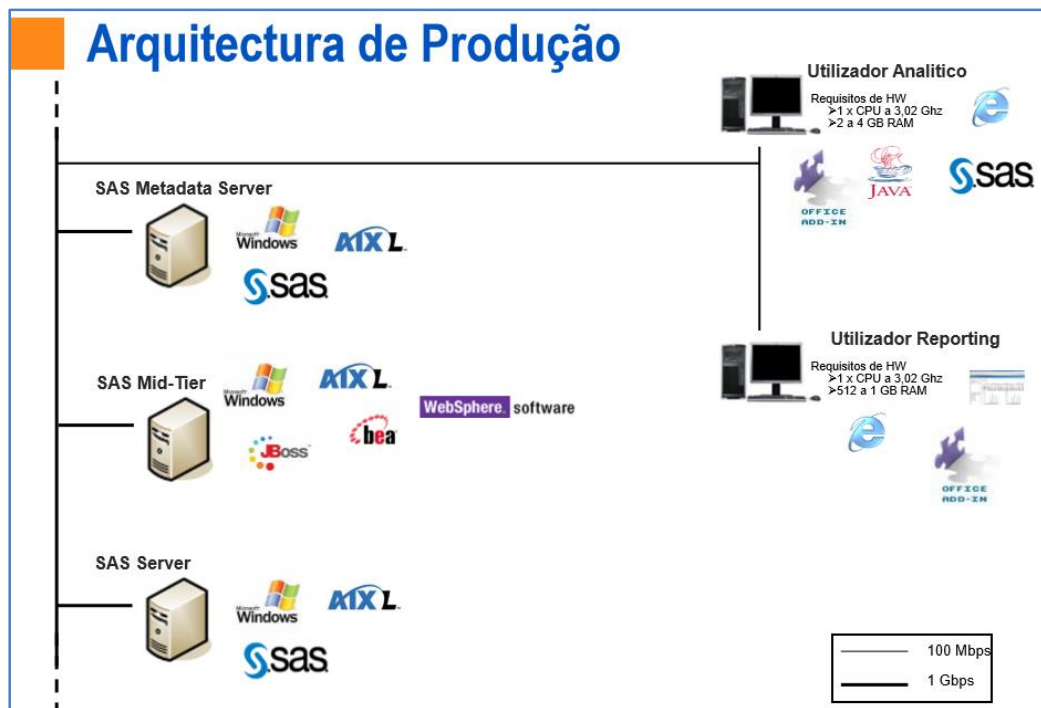
- SAS Enterprise Guide: módulo que possibilita a Visualização, Exploração, Manipulação, Selecção e Análise Estatística de Dados de uma forma simples e intuitiva (perfil de utilizador 1);
- SAS Add-in Office & OLAP: este módulo possibilita uma integração com ferramentas Office e o desenvolvimento de OLAP (perfil 1 de utilizador);
- SAS Enterprise Miner Desktop: módulo que permite o desenvolvimento de análises estatísticas e de DataMining avançadas (perfil 2 de utilizador).

2.2 Frameworks Integração

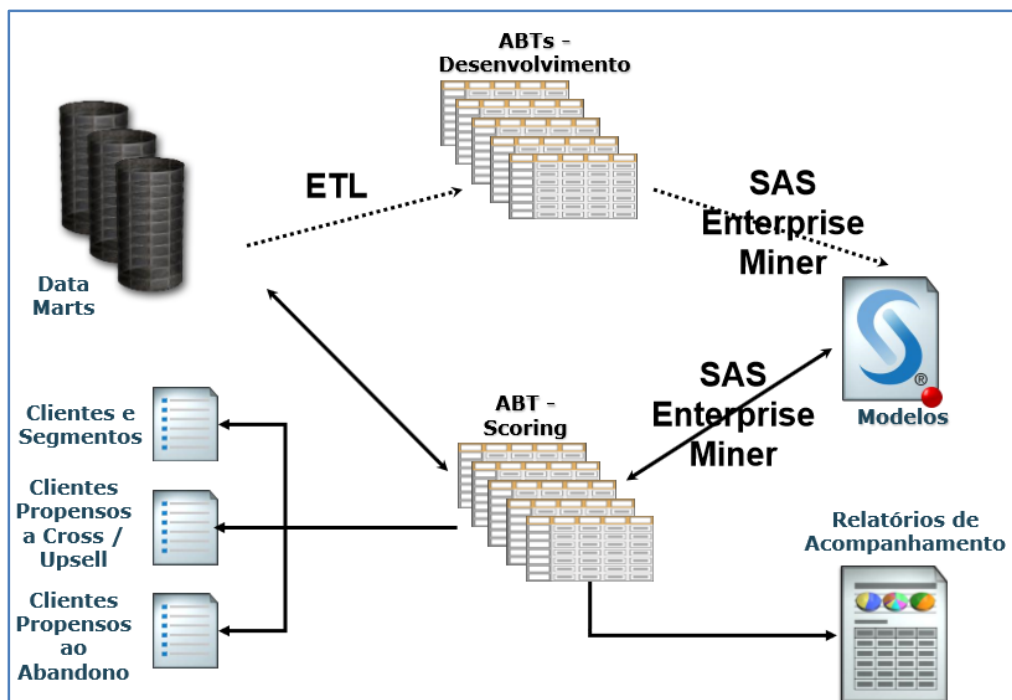
A plataforma SAS funciona com as seguintes frameworks de integração:

- JBoss Application Server
 - WebSphere Application Server
 - Oracle Web Logic Server Support for the Third Maintenance Release for SAS® 9.2 (TS2M3)
-

2.3 Design da Solução



2.4 Design de Integração



3. Gestão de recursos

3.1 Recursos Humanos

Project Manager

Gestor de projecto e responsável por todo o processo de comunicação entre equipas.

- Gestão de projectos e resolução de conflitos;

-
- Controlo de âmbito de intervenção e Estimativas;
 - Controlo de intervenção no Projecto

Analytical Consultant

Apoio na realização das análises, no desenvolvimento de Modelos e produção de relatórios. Apoio ainda a realização das actividades de Data Mining: selecção, exploração e modelagem.

- Conhecimentos analíticos e conhecimentos de negócio;
- Conhecimento sobre os dados disponíveis;

Information Management Consultant

Apoia o desenvolvimento dos processos de preparação de dados, na monitorização do decorrer dos processos de preparação de dados e relatórios e sua integração.

- Desenho de modelos de dados (Data Warehousing e Datamarts);
- Preparação de processos de manipulação de dados e integração;
- Experiência em programação e ferramentas.

Database System Administrator

Responsável por fazer a gestão, instalar, configurar, atualizar e monitorizar as BD's.

- Criar os mecanismos e testes de backup para garantir a recuperação dos dados no caso de falhas de hardware ou de sistema operativo
- Fazer a optimização de recursos disponíveis para garantir o correcto funcionamento ao longo do tempo
- Fazer a instalação e actualização das ferramentas utilizadas nas BD's;
- Saber os comandos e exclusivos de um sistema SGBD
- Verificar e garantir a integridade dos dados existentes na BD
- Garantir a performance (o máximo de desempenho para as consultas) de acesso aos dados
- Garantir o maior tempo possível de disponibilidade da BD
- Auxiliar a equipa de desenvolvimento e de testes para maximizar o uso e desempenho da BD

System Enginner

Conhecimentos no design, instalação, configuração e solução de problemas em sistemas de rede.

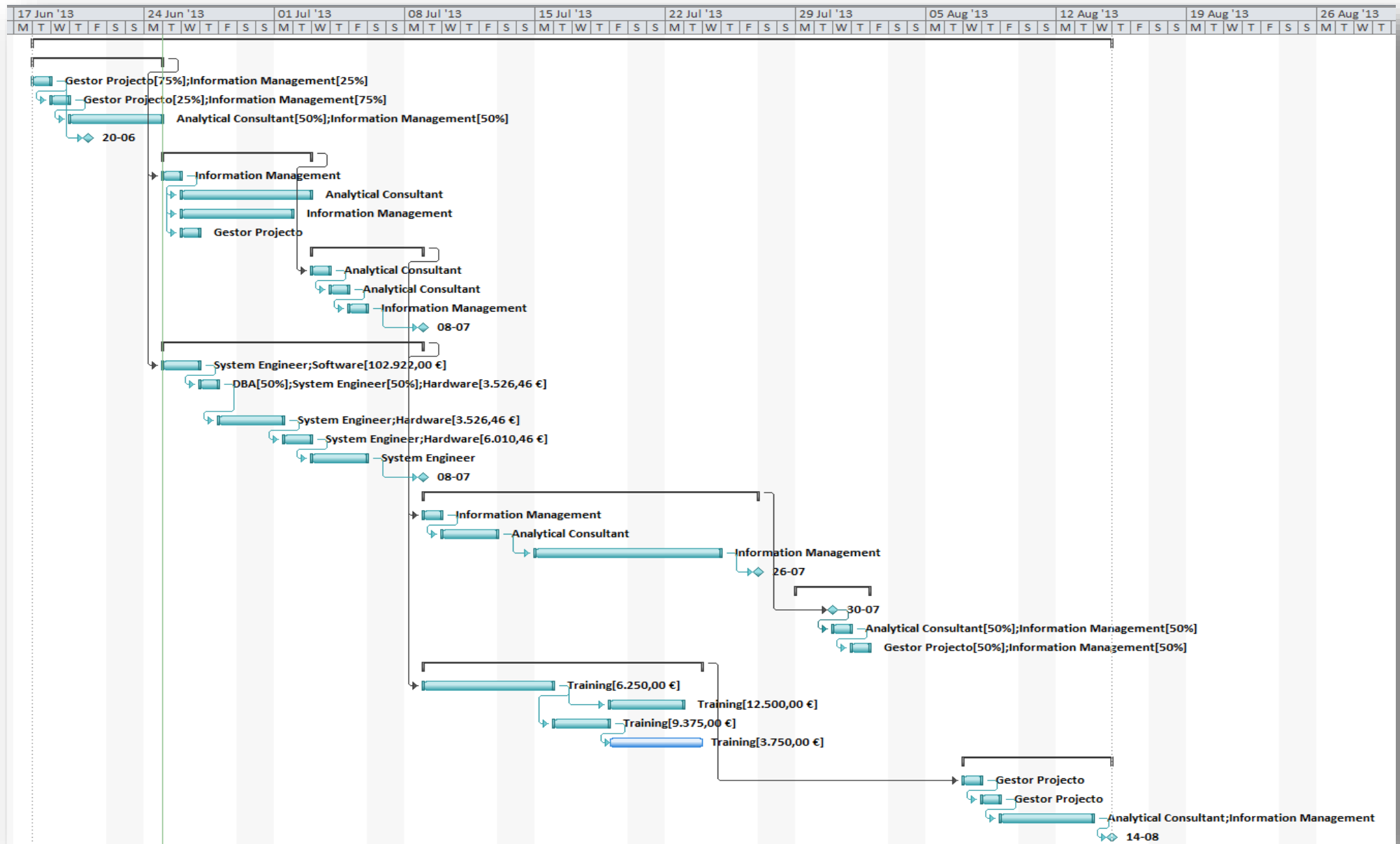
Recursos Internos

- SAS Enterprise Guide : 12 Utilizadores
 - SAS Add-in Office : 12 Utilizadores
 - Criação e exploração de OLAP: 12 Utilizadores
-

-
- Componentes servidor Portal para publicação e disponibilização corporativo dos resultados : 12 Utilizadores
 - Data Integration Studio para carregamento massivo e performance da informação: 2 Utilizadores
 - SAS Enterprise Miner Desktop: módulo que permite o desenvolvimento de análises estatísticas e de Data Mining avançadas (perfil 2 de utilizador): 2 Utilizadores
-

4. Diagrama de Gantt

	Task Name	Fixed Cost	Fixed Cost Accrual	Total Cost	Baseline	Variance	Actual	Remaining	WBS
1	GP HORECA	0,00 €	Prorated	163.988,38 €	0,00 €	163.988,38 €	0,00 €	163.988,38 €	1
2	Bussiness Understanding	0,00 €	Prorated	1.660,00 €	0,00 €	1.660,00 €	0,00 €	1.660,00 €	1.1
3	Determinar Objectivos do Negócio	0,00 €	Prorated	340,00 €	0,00 €	340,00 €	0,00 €	340,00 €	1.1.1
4	Determinar Objectivos de Data Mining	0,00 €	Prorated	300,00 €	0,00 €	300,00 €	0,00 €	300,00 €	1.1.2
5	Descrição da Solução	0,00 €	Prorated	780,00 €	0,00 €	780,00 €	0,00 €	780,00 €	1.1.3
6	Produzir o Plano do Projecto	0,00 €	Prorated	240,00 €	0,00 €	240,00 €	0,00 €	240,00 €	1.1.4
7	Data Understanding	0,00 €	Prorated	2.960,00 €	0,00 €	2.960,00 €	0,00 €	2.960,00 €	1.2
8	Recolha Inicial dos Dados	0,00 €	Prorated	280,00 €	0,00 €	280,00 €	0,00 €	280,00 €	1.2.1
9	Descrever os Dados	0,00 €	Prorated	1.200,00 €	0,00 €	1.200,00 €	0,00 €	1.200,00 €	1.2.2
10	Explorar os Dados	0,00 €	Prorated	1.120,00 €	0,00 €	1.120,00 €	0,00 €	1.120,00 €	1.2.3
11	Verificar a Qualidade dos Dados	0,00 €	Prorated	360,00 €	0,00 €	360,00 €	0,00 €	360,00 €	1.2.4
12	Data Preparation	0,00 €	Prorated	1.020,00 €	0,00 €	1.020,00 €	0,00 €	1.020,00 €	1.3
13	Seleccionar Dados	0,00 €	Prorated	240,00 €	0,00 €	240,00 €	0,00 €	240,00 €	1.3.1
14	Limpar Dados	0,00 €	Prorated	240,00 €	0,00 €	240,00 €	0,00 €	240,00 €	1.3.2
15	Construir Dados	0,00 €	Prorated	280,00 €	0,00 €	280,00 €	0,00 €	280,00 €	1.3.3
16	Integrar Dados	0,00 €	Prorated	260,00 €	0,00 €	260,00 €	0,00 €	260,00 €	1.3.4
17	System Instalation and Configuration	0,00 €	Prorated	119.013,38 €	0,00 €	119.013,38 €	0,00 €	119.013,38 €	1.4
18	Preparação de Ambientes	0,00 €	Prorated	103.530,00 €	0,00 €	103.530,00 €	0,00 €	103.530,00 €	1.4.1
19	Instalação e Configuração Metadata Server	0,00 €	Prorated	3.830,46 €	0,00 €	3.830,46 €	0,00 €	3.830,46 €	1.4.2
20	Instalação e Configuração Mid-Tier	0,00 €	Prorated	3.982,46 €	0,00 €	3.982,46 €	0,00 €	3.982,46 €	1.4.3
21	Instalação e Configuração Server	0,00 €	Prorated	6.466,46 €	0,00 €	6.466,46 €	0,00 €	6.466,46 €	1.4.4
22	Testes Unificados	0,00 €	Prorated	912,00 €	0,00 €	912,00 €	0,00 €	912,00 €	1.4.5
23	Documentação	0,00 €	Prorated	292,00 €	0,00 €	292,00 €	0,00 €	292,00 €	1.4.6
24	Modeling	0,00 €	Prorated	3.760,00 €	0,00 €	3.760,00 €	0,00 €	3.760,00 €	1.5
25	Seleccionar Técnicas de Modelação	0,00 €	Prorated	280,00 €	0,00 €	280,00 €	0,00 €	280,00 €	1.5.1
26	Criar uma Concepção de Testes	0,00 €	Prorated	720,00 €	0,00 €	720,00 €	0,00 €	720,00 €	1.5.2
27	Construir o Modelo	0,00 €	Prorated	2.240,00 €	0,00 €	2.240,00 €	0,00 €	2.240,00 €	1.5.3
28	Rever o Modelo	0,00 €	Prorated	520,00 €	0,00 €	520,00 €	0,00 €	520,00 €	1.5.4
29	Evaluation	0,00 €	Prorated	1.140,00 €	0,00 €	1.140,00 €	0,00 €	1.140,00 €	1.6
30	Avaliar os Resultados	0,00 €	Prorated	560,00 €	0,00 €	560,00 €	0,00 €	560,00 €	1.6.1
31	Rever o Processo	0,00 €	Prorated	260,00 €	0,00 €	260,00 €	0,00 €	260,00 €	1.6.2
32	Determinar os Próximos Passos	0,00 €	Prorated	320,00 €	0,00 €	320,00 €	0,00 €	320,00 €	1.6.3
33	Training	0,00 €	Prorated	31.875,00 €	0,00 €	31.875,00 €	0,00 €	31.875,00 €	1.7
34	Administration	0,00 €	Prorated	6.250,00 €	0,00 €	6.250,00 €	0,00 €	6.250,00 €	1.7.1
35	Data Manager	0,00 €	Prorated	12.500,00 €	0,00 €	12.500,00 €	0,00 €	12.500,00 €	1.7.2
36	Data Miner	0,00 €	Prorated	9.375,00 €	0,00 €	9.375,00 €	0,00 €	9.375,00 €	1.7.3
37	DI Developer	0,00 €	Prorated	3.750,00 €	0,00 €	3.750,00 €	0,00 €	3.750,00 €	1.7.4
38	Deployment	0,00 €	Prorated	2.560,00 €	0,00 €	2.560,00 €	0,00 €	2.560,00 €	1.8
39	Planear Avaliação de Resultados	0,00 €	Prorated	360,00 €	0,00 €	360,00 €	0,00 €	360,00 €	1.8.1
40	Planear Monitorização e Manutenção	0,00 €	Prorated	360,00 €	0,00 €	360,00 €	0,00 €	360,00 €	1.8.2
41	Produzir Relatório Final	0,00 €	Prorated	1.560,00 €	0,00 €	1.560,00 €	0,00 €	1.560,00 €	1.8.3
42	Rever o Projecto	0,00 €	Prorated	280,00 €	0,00 €	280,00 €	0,00 €	280,00 €	1.8.4



5. Matriz responsabilidades

Tem o objetivo de distribuição do trabalho, funções e responsabilidades das atividades dentro do projeto. Esta matriz identifica as responsabilidades de cada participante, com o intuito de reduzir conflitos.

Stakeholder Actividadez	Sponsor	IT Manager	Gestor de Projeto	Analytical Consultant	Information Management	Qualidade	DBA (Database System Administrator)	System Enginner
Plano	Aprova	Participa	Responsável	Participa	Participa	Exigida a revisão	Informado	Informado
Requisitos	Informado	Participa/Aprova	Participa	Responsável	Responsável	Exigida a revisão	Participa	Participa
Design	Informado	Participa	Aprova	Participa	Responsável	Exigida a revisão	Informado	Informado
Desenvolvimento	Informado	Participa	Aprova	Participa	Participa	Exigida a revisão	Participa	Informado
System Instalation and Configuration	Informado	Participa	Aprova	Participa	Participa	Exigida a revisão	Responsável	Responsável
Aquisições	Informado	Responsável	Aprova	Participa	Participa	Exigida a revisão	Informado	Informado
Integração	Informado	Participa	Participa	Responsável	Participa	Exigida a revisão	Participação	Participação
Teste	Aprova	Participa	Participa	Responsável	Participa	Exigida a revisão	Participa	Participa
Entregas	Aprova	Aprova	Participa	Participa	Participa	Informado	Informado	Informado
Plano de Comunicação	Aprova	Participa	Responsável	Participa	Participa	Exigida a revisão	Informado	Informado
Relatórios progresso	Informado	Informado	Responsável	Participa	Participa	Informado	Participa	Participa
Processo de alterações	Aprova	Informado	Responsável	Participa	Participa	Exigida a revisão	Participa	Participa

6. CBS (Costs Breakdown Structure)

Resource Name	Work	Cost
Training		31 875,00 €
Administration		6 250,00 €
Data Manager		12 500,00 €
Data Miner		9 375,00 €
DI Developer		3 750,00 €
Hardware		13 063,38 €
Instalação e Configuração Metadata Server		3 526,46 €
Instalação e Configuração Mid-Tier		3 526,46 €
Instalação e Configuração Server		6 010,46 €
Software		102 922,00 €
Preparação de Ambientes		102 922,00 €
Gestor Projecto	38 hrs	1 710,00 €
Determinar Objectivos do Negócio	6 hrs	270,00 €
Determinar Objectivos de Data Mining	2 hrs	90,00 €
Verificar a Qualidade dos Dados	8 hrs	360,00 €
Determinar os Próximos Passos	4 hrs	180,00 €
Planear Avaliação de Resultados	8 hrs	360,00 €
Planear Monitorização e Manutenção	8 hrs	360,00 €
Rever o Projecto	2 hrs	90,00 €
DBA	8 hrs	304,00 €
Instalação e Configuração Metadata Server	4 hrs	152,00 €
Documentação	4 hrs	152,00 €
System Engineer	68 hrs	2 584,00 €
Preparação de Ambientes	16 hrs	608,00 €
Instalação e Configuração Metadata Server	4 hrs	152,00 €
Instalação e Configuração Mid-Tier	12 hrs	456,00 €
Instalação e Configuração Server	12 hrs	456,00 €
Testes Unificados	24 hrs	912,00 €
Information Management	206 hrs	7 210,00 €
Determinar Objectivos do Negócio	2 hrs	70,00 €
Determinar Objectivos de Data Mining	6 hrs	210,00 €
Descrição da Solução	12 hrs	420,00 €
Recolha Inicial dos Dados	8 hrs	280,00 €
Explorar os Dados	32 hrs	1 120,00 €

Construir Dados	8 hrs	280,00 €
Integrar Dados	4 hrs	140,00 €
Documentação	4 hrs	140,00 €
Seleccionar Técnicas de Modelação	8 hrs	280,00 €
Construir o Modelo	64 hrs	2 240,00 €
Rever o Modelo	8 hrs	280,00 €
Avaliar os Resultados	16 hrs	560,00 €
Rever o Processo	4 hrs	140,00 €
Determinar os Próximos Passos	4 hrs	140,00 €
Produzir Relatório Final	24 hrs	840,00 €
Rever o Projecto	2 hrs	70,00 €
Analytical Consultant	144 hrs	4 320,00 €
Descrição da Solução	12 hrs	360,00 €
Produzir o Plano do Projecto	8 hrs	240,00 €
Descrever os Dados	40 hrs	1 200,00 €
Seleccionar Dados	8 hrs	240,00 €
Limpar Dados	8 hrs	240,00 €
Integrar Dados	4 hrs	120,00 €
Criar uma Concepção de Testes	24 hrs	720,00 €
Rever o Modelo	8 hrs	240,00 €
Rever o Processo	4 hrs	120,00 €
Produzir Relatório Final	24 hrs	720,00 €
Rever o Projecto	4 hrs	120,00 €

7. Proposta final

Hardware	Qtd	Unit.	Total
Metadata Server	1	3 526,46 €	3 526,46 €
Mid-Tier	1	3 526,46 €	3 526,46 €
Sever	1	6 010,46 €	6 010,46 €
			13 063,38 €

Software	Qtd	Unit.	Total
Windows Server 2008 R2 x64	3	200,00 €	600,00 €
SAS (Software)	1	73 522,00 €	73 522,00 €
Licenciamento - Perfil 1	10	2 250,00 €	22 500,00 €
Licenciamento - Perfil 2	2	3 150,00 €	6 300,00 €
			102 922,00 €

Formação	Qtd	Unit.	Total
SAS® Plataform Administrator Fast Track	2	3 125,00 €	6 250,00 €
SAS® Enterprise Guide® I: Querying and Reporting	5	1 250,00 €	6 250,00 €
SAS® Enterprise Guide® 2: Advanced Tasks and Querying	5	1 250,00 €	6 250,00 €
Applied Analytics Using SAS® Enterprise Miner™	3	1 875,00 €	5 625,00 €
Predictive Modeling Using Logistic Regression	3	1 250,00 €	3 750,00 €
SAS® Data Integration Studio 1: Essentials	2	1 875,00 €	3 750,00 €
			31 875,00 €

Implementação	Qtd	Unit.	Total
Bussiness Understanding	1	1 660,00 €	1 660,00 €
Data Understanding	1	2 960,00 €	2 960,00 €
Data Preparation	1	1 020,00 €	1 020,00 €
System Instalation and Configuration	1	3 028,00 €	3 028,00 €
Modeling	1	3 760,00 €	3 760,00 €
Evaluation	1	1 140,00 €	1 140,00 €
Deployment	1	2 560,00 €	2 560,00 €
			16 128,00 €