



• C •

FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Departamento de Engenharia Electrotécnica e de Computadores
Mestrado em Tecnologia de Informação Visual

**Framework de simulação de modelos de scoring para
melhorias de eficiência na gestão de leads**

Filipe Agante Simões

Dissertação realizada sob orientação de:
Professor Doutor Nuno Miguel Mendonça da Silva Gonçalves

Coimbra, Julho de 2012

Agradecimentos

Queria agradecer ao Professor Doutor Nuno Gonçalves, meu orientador, que sempre se disponibilizou para me auxiliar e nas alturas mais importantes sempre contribuiu de forma decisiva para a realização desta dissertação.

Uma palavra especial para as pessoas que me ajudaram durante os meus estudos, nomeadamente na concretização do mesmo, com a elaboração desta dissertação. Em particular queria agradecer aos meus pais, que sempre me ajudaram e possibilitaram a concretização dos meus estudos, à minha irmã por ser a minha preferida, à minha namorada Mariana que sempre me acompanhou e apoiou nos dias menos bons durante a elaboração do trabalho e a todas as pessoas que de uma ou de outra forma me ajudaram durante a dissertação

Resumo

Neste trabalho analisamos diferentes modelos de *scoring* de forma a melhorar a eficiência na gestão de *leads*. A área de estudo é a restauração, nomeadamente os restaurantes em Coimbra e nos seus concelhos periféricos. Para isso, construímos uma plataforma de simulação para simular a selecção de prestadores de serviço e encaminhamento de clientes que, registando-se na plataforma através de *microsites*, pretendem ir a um restaurante com determinadas características. Para este estudo ser possível, criou-se ainda, um cenário onde estão presentes os clientes e os restaurantes que se encontram registados na plataforma.

Um aspecto diferenciador nos modelos de *scoring* desenvolvidos e testados neste trabalho, é que analisam os restaurantes e não os clientes. A análise tenta compreender quais são os atributos dos restaurantes e qual a sua importância na decisão do cliente ir ao restaurante.

Desta forma, o objectivo principal deste estudo é encontrar modelos de *scoring* que aumentem a taxa de conversão dos restaurantes.

Para o estudo destes modelos foram produzidos três tipos de testes. O primeiro teste tem como objectivo identificar os modelos onde a número de conversões é mais elevado e daí concluir quais os atributos que contribuem de forma positiva para o aumento da taxa de conversão. No segundo teste foi analisado o comportamento de novos restaurantes que ao longo do tempo vão sendo inseridos na plataforma. Por fim no último teste foram analisados três restaurantes com diferentes taxas de conversão (baixa, média e alta), de modo a estudar o percurso das suas taxas de conversão ao longo da simulação.

Após os testes e pela análise dos resultados obtidos, ficou demonstrado que cada atributo determina um resultado diferente na taxa de conversão, ou seja, os atributos influenciam de forma díspar o número de conversões. Podemos ainda concluir, que os modelos de *scoring* podem objectivamente provocar um aumento do número de conversões. Este facto demonstra assim que, os modelos de *scoring* melhoram a eficiência na gestão de *leads*.

Palavras-Chave

Lead, Modelo de *Scoring*, Plataforma, Simulador, Clientes, Prestadores de Serviço, Taxa de Conversão

Abstract

This study analyzes different scoring models to increase efficiency in lead management. The area of study is the restaurant industry, specifically restaurants in Coimbra and its peripheral municipalities. For these, we constructed a simulator platform to simulate the selection and forward of clients, who intend to go to a restaurant with specific characteristics. To make this study possible, we created one virtual scenario, with clients and restaurants which are both registered at the platform.

The scoring models developed and tested in this study analyze the restaurants and not the clients. The analysis attempts to understand which are the main attributes of the restaurants and what are their importance in the client's decision to go to the restaurant.

The main objective of this study is to find scoring models that increase the restaurants conversion rate.

For the study of these models we developed three types of test. The first test has the purpose to identify the models where the number of conversions is higher, and from that, conclude which attributes contribute positively to the increase of the conversion rate. The second test has been analyzed the behavior of new restaurants along the time. Finally the last test analyzed three restaurants with different conversion rates (low, medium and high), in order to study the progress of their conversion rates throughout the simulation.

From the results, it was demonstrated that each attribute determines a different result in conversion rate, ie, the attributes affect differently the number of conversions. We can also conclude that the scoring models can cause an increase in the number of conversions. This demonstrates that in fact, the scoring models can improve efficiency in lead management.

Keywords

Lead, Scoring Model, Platform, Simulator, Providers, Customers, Conversion Rate

Conteúdo

Agradecimentos	ii
Resumo	iii
Palavras-Chave	iii
Abstract	iv
Keywords	iv
Capítulo 1	1
Introdução	1
1.1 Evolução dos tempos.....	1
1.2 Meios de geração de novos clientes	2
1.3 Lead Management.....	4
1.4 Descrição do Problema.....	5
1.5 Objectivos.....	6
1.6 Estrutura da Dissertação	7
Capítulo 2	8
O aparecimento dos modelos de <i>scoring</i>	8
2.1 Diferentes aplicações	8
2.2 Modelos de <i>scoring</i> na gestão de leads	9
Capítulo 3	11
Plataforma de Simulação	11
3.1. Vantagens de existir uma Plataforma de Simulação.....	11
3.2 Ambiente de desenvolvimento	11
3.3 Arquitectura da Plataforma de Simulação	12
3.4 Funcionamento da Plataforma de Simulação	12
3.4.1 Customers.....	16
3.4.2 Providers.....	17
3.4.3 Plataforma	18
3.4.4 Simulador	19
Capítulo 4	22
Cenário	22
4.1 Construção do Cenário	25
4.1.1 Inserção esporádica de novos restaurantes.....	27
Capítulo 5	29
Modelos de <i>scoring</i> e bonificações das taxas de conversão.....	29
5.1 Modelos de <i>scoring</i>	29
5.1.1 Modelos Simples	29
5.1.1 Modelos Complexos	30
5.2 Bonificação da Taxa de Conversão	33
5.2.1 Low Cost	33
5.2.2 Quality vs Price	33
5.2.3 High Quality.....	34
Capítulo 6	35

Testes e Resultados	35
6.1 – Testes.....	35
6.2 Resultados	36
Capítulo 7	45
Conclusão e Trabalho Futuro	45
7.1 Conclusão	45
7.2 Trabalho Futuro.....	47
Bibliografia	48
ANEXOS	49

Lista de Figuras

Fig. 1 – Figura representativa de todo o processo, desde o registo do cliente até à sua ida ao restaurante.	4
Fig. 2 – Diagrama de Arquitectura.....	13
Fig. 3 – Diagrama de actividades.....	14
Fig. 4 – Diagrama de Casos de Uso de cada um dos agentes	14
Fig. 5 – Imagem da Framework a destacar as principais áreas	15
Fig. 6 – Diagrama de Interação – Registo no <i>site</i>	16
Fig. 7 – Diagrama de actividade do cliente.....	17
Fig. 8 – Diagrama de actividades da Plataforma.....	19
Fig. 9 – Esquema que representa as CR dos restaurantes seleccionados a e probabilidade de não existir conversão.....	20
Fig. 10 – Diagrama de actividade do Simulador.....	21
Fig. 11 – Gráficos que elucidam a forma como o Score evolui com o aumento da CR e do NC.	32
Fig. 12 – Gráfico com as taxas de conversão nas 24 combinações.....	37
Fig. 13 – Gráfico da taxa de conversão média dos novos restaurantes pelo modelo <i>Random</i>	39
Fig. 14 - Gráfico da taxa de conversão média dos novos restaurantes pelo modelo CR	39
Fig. 15 - Gráfico da taxa de conversão média dos novos restaurantes pelo modelo BCG-1	40
Fig. 16 – Gráfico da taxa de conversão de R1 no modelo <i>Random</i>	42
Fig. 17 - Gráfico da taxa de conversão de R1 no modelo <i>CR</i>	42
Fig. 18 - Gráfico da taxa de conversão de R1 no modelo BCG 1	42
Fig. 19 - Gráfico da taxa de conversão de R2 no modelo <i>Random</i>	43
Fig. 20 - Gráfico da taxa de conversão de R2 no modelo <i>CR</i>	43
Fig. 21 - Gráfico da taxa de conversão de R2 no modelo BCG 1	43
Fig. 22 - Gráfico da taxa de conversão de R3 no modelo <i>Random</i>	44
Fig. 23 - Gráfico da taxa de conversão de R3 no modelo <i>CR</i>	44
Fig. 24 - Gráfico da taxa de conversão de R3 no modelo BCG 1	44

Lista de Tabelas

Tabela 1 - Variáveis relativas ao Simulador	23
Tabela 2 – Variáveis relativas ao Clientes e Restaurantes	24
Tabela 3 – Matriz de pesos para cálculo do Score	32
Tabela 4 – Matriz com duas entradas (<i>Distance</i> e <i>Price</i>) para calcular o valor da bonificação ...	33
Tabela 5 – Matriz com quatro entradas (<i>CR</i> , <i>RS</i> , <i>Price</i> e <i>Distance</i>) para calcular o valor da bonificação	34
Tabela 6 - Matriz com duas entradas (<i>Price</i> e <i>CR</i>) para calcular o valor da bonificação	34
Tabela 7 – Quadro resumo dos resultados do teste do tipo 1	36
Tabela 8 – Taxa de conversão média por tipo de bonificação	38
Tabela 9 - Quadro resumo dos resultados dos novos restaurantes.....	38
Tabela 10 – Quadro resumo do nº de conversões dos novos restaurantes	40
Tabela 11 - Quadro resumo do nº de selecções dos novos restaurantes.....	41
Tabela 12 - Quadro resumo do nº de conversões dos 500 restaurantes existentes.....	41
Tabela 13 - Quadro resumo do nº de selecções dos 500 restaurantes existentes	41
Tabela 14 – Tabela completa dos resultados referentes ao testes do Tipo 1	50

Capítulo 1

Introdução

1.1 Evolução dos tempos

Um dos principais factores a ter em conta no mundo empresarial é a competitividade. A posição em que uma empresa se situa em comparação com as restantes, que competem pelo mesmo objectivo, resulta no sucesso ou no insucesso do negócio.

O modo como os clientes olham para cada empresa é determinante na hora da decisão. São por isso, necessários, mecanismos de sedução para aproximar os clientes às empresas ou aos seus serviços.

Apesar de hoje em dia o conhecimento no ambiente empresarial ser elevado, continuam todos os anos a existir empresas que fecham as suas portas. Por outro lado, continua a haver pessoas que prosseguem o seu sonho de abrir a sua própria empresa. A razão mais premente para o encerramento de negócios é a perda de lucros ao longo de vários anos sem que se consiga inverter a tendência. O decréscimo do volume de negócios pode ter géneses completamente distintas, sendo que uma das principais razões nos últimos anos tem sido o aumento dos meios da internet e das infra-estruturas que lhe são inerentes.

Hoje, quase todos os negócios estão directa ou indirectamente conectados com algum elemento do mundo “*web*”. O aumento de competitividade, trazido pelo aumento do comércio *online* (Li et al., 2010), bem como a partilha de informação são cada vez mais observados, factos que conduzem a que os segredos no universo das empresas sejam cada vez menos exclusivos.

Por estas razões, o modo de contacto entre clientes e fornecedores de serviço mudou substancialmente. Os processos usados pelo marketing para angariar e manter clientes sofreram modificações, nomeadamente na gestão da relação com os clientes (CRM – *Customer Relationship Management*), sendo este um dos métodos mais usados nestes processos (Srivastava, Shervani, & Fahey, 1999). O CRM era tradicionalmente reconhecido como um modelo de relacionamento B2B (business2business) ou B2C (business2customer), baseado no conhecimento dos hábitos dos clientes pelo seu historial de compras. Contudo, as mudanças

recentes levaram a que este modelo fosse reconhecido como um método para reter os clientes e otimizar as vendas e os lucros, monitorizando todas as suas interações. É assim, criado um passaporte para cada cliente, onde está presente a informação relevante, nomeadamente dados pessoais, contactos feitos e historial de compras. Com esta informação privilegiada de cada cliente, é possível desenvolver um relacionamento de longo prazo (Burkey & Breakfield, 2001).

No marketing o sucesso traduz-se muita vez pelo “agarrar” das oportunidades sempre que estas surjam. Um potencial novo cliente, designado de *lead* em marketing, é uma oportunidade de realizar mais um negócio, todavia para isso é necessário atrair a *lead* para o “nosso lado” ao invés de permitir que a concorrência a capte.

1.2 Meios de geração de novos clientes

A indústria da publicidade *online* teve início nos anos 90, tendo como primeiro impulsionador o aparecimento das *dot-com*, readquirindo um novo ímpeto com o modelo *Web 2.0*. Com este novo modo de comunicação, novas formas de contacto com os clientes sobrepuseram-se aos formatos de aproximação usados até então. As redes sociais, *newsletters*, blogs, vídeos e *microsites* ganharam um papel fundamental no contacto e meio de interacção com os mesmos.

Após a mudança no modo de comunicação, e uma vez que a maioria dos processos foi transportado para a web, é necessário capturar o maior número de pessoas, para os sítios onde os serviços das empresas são divulgados. Para isso, novos métodos foram e continuam a ser criados, de modo a capturar o máximo de tráfego possível.

Os métodos mais frequentes são feitos através dos motores de busca, nomeadamente no *Google*, *Yahoo*, *Bing* ou no Sapo no caso português. Os motores de busca são quase sempre, o início para um cliente que procura um determinado bem ou serviço através da internet e devido a isso, uma boa colocação nos motores de busca significa, por si, um aumento substancial da probabilidade de o cliente clicar num determinado *site*.

Devido à importância dos motores de busca, criou-se um conjunto de estratégias de forma a otimizar os *sites*, tendo em conta os algoritmos de ordenamento dos motores de busca. Os algoritmos foram-se tornando cada vez mais completos, analisando diversas variáveis, para o resultado ser o melhor possível, ou seja, que os conteúdos exibidos pelas páginas sejam o que realmente o utilizador procura.

Um das formas que as empresas têm encontrado para conseguir focar as suas novas campanhas e serviços na web, nomeadamente nos motores de busca, tem sido a construção de *microsites*.

Os *microsites* têm assumido um papel cada vez mais importante neste objectivo. Este novo modelo de *site* permite focar a campanha num determinado produto, serviço ou segmento de cliente alvo. As grandes companhias constroem um *microsite* para cada nova campanha que surge, com *urls* diferentes do principal *site* da marca. Funcionando independentemente, o *microsite* pode ser mais facilmente optimizado para os motores de busca, devido à possibilidade de estar evidenciado num conjunto de palavras (*keywords*) ou frase e assim permite um acesso mais fácil aos consumidores. O seu tamanho reduzido traz uma maior agilidade e rapidez no ajustar do mesmo aos desejos dos consumidores.

Devido a todas estas vantagens, muitos dos *microsites* são especializados na obtenção de *leads*, que posteriormente são recebidas por sistemas de *Lead Management* (secção seguinte) e distribuídas pelos prestadores de serviço. Podemos definir uma *lead* como sendo um potencial cliente identificado através dos métodos de angariação já descritos.

Com o aumento substancial de pessoas que utilizam a internet, os potenciais clientes também aumentaram abruptamente. Surgiram assim empresas de marketing especializadas tanto na geração de *leads* como na sua gestão e estudo. Podemos então identificar os protagonistas deste universo: os prestadores de serviços cooperam com empresas que utilizam estes processos, de forma a lhes indicar as *leads* que estão interessadas nos seus serviços (figura 1).

Com a especialização destas empresas na gestão de *leads* surgiu uma prática denominada *Lead Management*, que será descrita na secção seguinte.

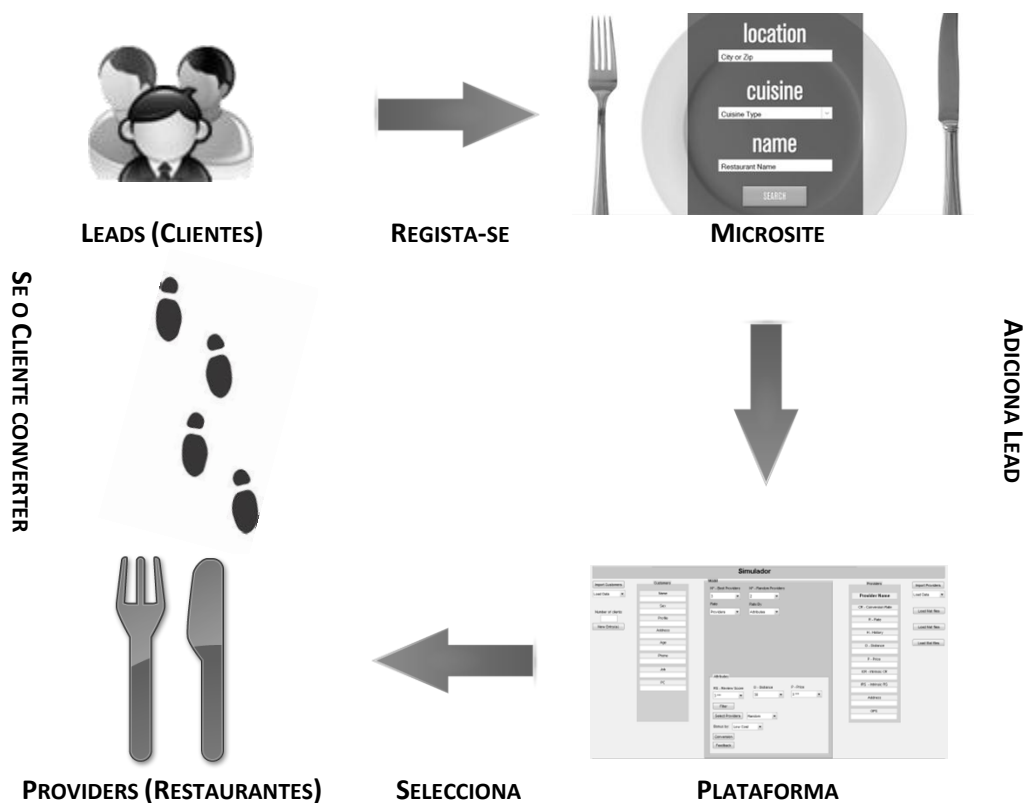


Fig. 1 – Figura representativa de todo o processo, desde o registo do cliente até à sua ida ao restaurante.

1.3 Lead Management

Como vimos, Lead Management (LM) é uma prática usada desde a captura de clientes até à concretização de um negócio. É por isso um processo complexo que pode ser decomposto em várias etapas. A primeira etapa do LM é o Lead Generation (LG), processo de obtenção de potenciais clientes através dos meios de publicidade já referidos.

A qualidade de cada *lead* obtida por cada fonte de publicidade é distinta, podendo ser medida pelo tipo de informação recolhida, bem como, pela quantidade de informação. É assim, vantajoso recolher o máximo de informação possível. Uma das formas mais recorrentes é o uso de *banners* publicitários *online*, que nos guiam até formulários. Este método de recolha de informação, é um dos mais comuns, uma vez que se pode obter um vasto conjunto de informação numa única interação com a *lead*.

Após a obtenção das *leads* é imprescindível saber como usar essa fonte de potenciais clientes. Sem esta gestão, é infrutífero o trabalho efectuado na aquisição dos *leads*. Como vimos anteriormente nem todos os contactos partem do mesmo patamar e por isso deverá ser feita uma selecção dos *leads* com maiores probabilidades de sucesso. No entanto é vantajoso criar um

conjunto de parâmetros mínimos estritamente necessários para que a *lead* possa ser qualificada. Apesar deste facto, existem *leads* qualificadas em que a informação é tão escassa, que impossibilita qualquer estudo sobre elas. Nestes casos a metodologia LM executa o processo de *Lead Nurturing* (LN), também referido como incubação de clientes, sobre as *leads*.

O LN é um processo que consiste no aumentar dos esforços de contacto com as *leads*, de modo a prepará-las para terem maiores probabilidades de serem futuros clientes, ou seja, são tentativas de aumentar a quantidade de informação para o passaporte de cada cliente e com isso reforçar a relação entre cliente e empresa. As *leads* que estão preparadas para ser qualificadas seguem para um processo de ordenação denominado *Lead Scoring*.

O *Lead Scoring* não é mais do que uma ordenação obtida pelo estudo dos atributos de cada *lead* compilado num único valor. Este score avalia a probabilidade da *lead* ser um potencial cliente. A ordenação é feita por modelos de *scoring*, analisando várias variáveis em simultâneo tendo em conta o target do mercado. Além da informação adquirida de cada *lead* necessária para a ordenação, existe informação que não é relevante para a mesma ordenação mas relevante noutras áreas.

A informação complementar, como por exemplo a localização geográfica, é vital para o sector de vendas. Imaginemos uma empresa “X” que possui vários pontos de venda no território nacional. Neste caso, a localização do cliente é de grande relevância para o sector de vendas incumbir a transacção ao seu ponto de venda mais próximo do cliente.

Em suma, a existência de Lead Management advém da necessidade de as empresas pretenderem aumentar a eficácia das vendas e, conseqüentemente, o lucro gerado.

Apesar das vantagens apresentadas, o Lead Management não tem resultados imediatos e tem os seus custos dependentes da complexidade do processo. Existe uma relação custo-benefício; o processo apenas será lucrativo se de facto aumentar o número de vendas.

1.4 Descrição do Problema

Quando uma *lead* é recebida, a probabilidade de concretização do negócio é tanto maior quanto melhor for a qualidade do ou dos prestadores de serviço. Existe, assim, uma relação entre as características dos prestadores de serviço e o grau de concretização das *leads* uma vez que as pessoas procuram determinadas especificidades e se o seu pedido não for satisfeito a probabilidade da *lead* concretizar é consideravelmente mais pequena. Deste modo, é fundamental desenvolver algoritmos que possam indicar quais os melhores prestadores de serviço para cada cliente. Esta solução pode ser encontrada por um agente intermediador que indique aos

prestadores de serviços quais as pessoas que podem estar interessadas no seu serviço. A *framework* desenvolvida decorre do trabalho aqui apresentado, desempenhando assim a função de seleccionar os prestadores de serviço e encaminhar-lhes as *leads* de potenciais clientes.

Existem diversas áreas de negócio onde esta solução pode ser adoptada: hotelaria, restauração, imobiliário e indústria mas o estudo incide sobre o sector da restauração, nomeadamente restaurantes. A melhor estratégia de abordagem ao problema é começar por perceber quais as variáveis que melhor caracterizam as *leads* e os restaurantes. Nas *leads* os dados tentam reflectir o seu comportamento na vida real, e as variáveis dos restaurantes reproduzem a oferta que possuem, bem como a localização e o seu historial. Os algoritmos de *scoring* entram, assim, com as características dos restaurantes e clientes, seleccionando os melhores restaurantes para cada cliente de forma que a probabilidade de conversão seja a maior possível. O estudo não calcula apenas a probabilidade de conversão mas simula também se o cliente vai efectivamente ao restaurante. Deste modo, é possível saber com exactidão quais os modelos que provocam um aumento na percentagem de concretização dos clientes. Pode-se ainda retirar dos modelos com taxas de conversão mais elevadas, quais são os atributos em concreto dos restaurantes, que levam o cliente a optar por este em detrimento doutro.

1.5 Objectivos

Em diversas áreas de negócio existem empresas que têm como função servirem de intermediários entres clientes e prestadores de serviço, para isso, utilizam processos de angariação e gestão de *leads*. A eficiência dos processos resulta ou não na eficácia de todo o modelo. De forma a aumentar a eficiência é, pois, essencial a existência de plataformas que permitam fazer toda a gestão necessária, quer na recepção das *leads* e *providers*, quer na selecção destes.

A plataforma de simulação desenvolvida neste trabalho é uma dessas ferramentas que nos vai permitir estudar e antecipar os possíveis comportamentos de clientes que procuram restaurantes especificamente na zona de Coimbra. A análise passa por compreender quais os atributos dos restaurantes e de que maneira são importantes na decisão do cliente ir efectivamente ao restaurante, através de modelos de *scoring*. O objectivo principal do trabalho é o desenvolvimento de modelos de *scoring* que aumentem o número de clientes que vão ao restaurante. Este objectivo implica também o desenvolvimento de uma plataforma dinâmica, onde é possível estudar um ou mais cenários com *providers* (restaurantes) e *customers* (clientes) de uma forma realista, aplicando os diversos modelos de *scoring* desenvolvidos.

O trabalho desenvolvido nesta dissertação tem certamente um contributo positivo para o estudo de modelos de *scoring* no processo de selecção de restaurantes e na bibliografia pouco vasta nesta área. A plataforma desenvolvida, possibilita ainda o estudo de outros modelos de *scoring* para o encaminhamento de *leads*, não só na área de restauração. Permite o estudo noutras áreas onde os processos sejam idênticos, como por exemplo na hospedagem.

1.6 Estrutura da Dissertação

A dissertação está dividida em seis capítulos, o primeiro desses, a “Introdução” que termina nesta secção com a descrição da estrutura do trabalho. O segundo capítulo, “O aparecimentos dos modelos de *scoring*”, é sobre o ponto central deste estudo, ou seja, os modelos de *scoring*, sendo enunciadas as áreas onde eram mais frequentemente usados e a sua expansão para outras áreas de negócio. Neste capítulo é ainda fundamentada a vantagem de existir modelos de *scoring* na gestão de *leads*.

No terceiro capítulo é explicada a razão do benefício de uma plataforma de simulação neste tipo de estudos, é ainda descrito o ambiente em que foi desenvolvida, bem como, a sua arquitectura. São ainda analisados e enumerados os principais agentes que a *Framework* comporta, e descrito o funcionamento da *Framework*.

O capítulo quatro centra-se no estudo do Plataforma de Simulação, sendo detalhada a forma como foram gerados os restaurantes e os clientes, nomeadamente como foi criado cada atributo e o seu significado no estudo.

No quinto capítulo vão ser enunciados e descritos os modelos de *scoring* e as bonificações nas taxas de conversão, ambos usados nas simulações.

No penúltimo capítulo serão apresentados os diferentes tipos de teste e os resultados decorrentes das várias simulações perante os diversos modelos de *scoring*.

No último capítulo é apresentado, uma conclusão de todo o trabalho, e os possíveis desenvolvimentos que poderão ser feitos no futuro.

Capítulo 2

O aparecimento dos modelos de *scoring*

2.1 Diferentes aplicações

As técnicas de modelo de *scoring* mais elaboradas começaram a ser usadas em sistemas de crédito, de forma a ajudar na decisão de conceder ou não o crédito ao cliente (Thomas, 2000). Thomas usa dois tipos de *scoring*: um *scoring* de crédito e um *scoring* de comportamento. Ambas as técnicas ajudam na decisão, embora de forma distinta. Na primeira técnica o modelo de crédito ajuda o banco a tomar a decisão de se deve ou não atribuir o crédito a um novo cliente. Na segunda técnica, o modelo de *scoring* de comportamento dá a informação se o crédito deve ser concedido a um cliente que já tem um histórico no banco, ou seja, o modelo usa informação recolhida para prever o comportamento do cliente.

O comportamento por norma é medido pelos dados pessoais: estado civil, rendimentos, despesas, contas bancárias, valor do património, sexo, entre outros. Todos estes factores são analisados e compilados num único *score*, se este atingir um determinado valor, existe assim o aval do banco para o crédito ser concedido.

Com o decorrer do tempo, estas técnicas, usadas na maioria por instituições de crédito expandiram-se por outra áreas, onde a previsão pode trazer vários benefícios, nomeadamente na redução de riscos e na diminuição de custos.

O marketing foi um dos sectores que adoptou estas técnicas com sucesso (Chong, 2010). Nesta área é especialmente importante prever os comportamentos das pessoas antes de se tomarem decisões de forma a evitar possíveis erros, que sem estas técnicas são dificilmente detectáveis. Estes procedimentos aumentam assim a eficiência dos procedimentos, ajudando na concretização dos objectivos das empresas. Estes estudos podem ser efectuados em campanhas de angariação de clientes, prevendo as necessidades dos clientes ou identificar novas oportunidades de negócio em nichos de mercados que outrora não tenham sido explorados.

Os modelos de *scoring* mais simplistas definem um conjunto de parâmetros iniciais e a importância de cada um deles. Posteriormente verificam para cada ponto se a lead atinge os critérios mínimos, se conseguir ultrapassar os critérios, é-lhe atribuído um valor. O score final é a somatório de todos os pontos conseguidos em cada parâmetro definido. Quanto maior for o score final, maior é a qualidade da lead e maior será a probabilidade desta se tornar num futuro cliente.

Os modelos mais complexos são os que estudam o comportamento de pessoas ou entidades. Uma vez que os comportamentos humanos têm um elevado grau de complexidade e de imprevisibilidade, a construção destes modelos pode ser abordada de forma muito diferente, apesar das diversas interpretações poderem ser correctas.

Os modelos de comportamento tiveram a mesma origem do marketing directo, nomeadamente o *database marketing* (Chong, 2010). O *database marketing* diferencia-se das concessões de crédito porque neste caso é preciso construir uma relação complexa com base na possível rentabilidade que um cliente possa trazer. Nos sistemas bancários a decisão tem apenas duas respostas possíveis, a concessão do crédito é dada ou não, apesar de que um erro nesta decisão poder provocar percas significativas.

Recentemente as técnicas de *data mining* têm sido utilizadas no marketing e no sector de vendas das empresas (Berry & Linoff, 1997; Adriaans & Zantinge, 1996; Berson, Smith, & Thearling, 1999).

O *data mining* constrói modelos usando informação real recolhida ao longo do tempo depositada numa base de dados. Através de técnicas de estudo a partir desses dados pode-se prever os comportamentos dos clientes. A chave destes algoritmos de predição é encontrar respostas quando não existe informação nalguns campos da base de dados. Este processo é muitas vezes denominado de *scoring*, porque os modelos de predição são frequentemente usados, quando se tem de atribuir um valor a uma característica, (Berry & Linoff, 1997). O valor numérico reflecte cada campo da base de dados e indica a probabilidade de o cliente ter um determinado comportamento (Berson, Smith, & Thearling, 1999).

2.2 Modelos de *scoring* na gestão de leads

Com a entrada dos modelos de *scoring* noutras áreas de acção, nomeadamente no marketing, o Lead Management adoptou estas técnicas de forma a aumentar a eficiência dos seus processos.

Desta forma como vimos no Capítulo 1, o processo de Lead Management usa algoritmos de *scoring* de forma a ordenar e identificar as leads (clientes ou empresas) em que o grau de certeza de um novo negócio seja maior. Os modelos de *scoring* podem assim trazer largos benefícios na aquisição de novos clientes para as empresas e ser um dos seus maiores aliados, sabendo-se que angariar um novo cliente é muito mais dispendioso do que manter um cliente (Kotler, Bowen, & Makens, 2006).

Neste estudo os modelos de *scoring* vão ser utilizados sobre os prestadores de serviço e não sobre os clientes como é mais usual. Vão ser usados modelos de *scoring* de comportamento, uma vez que existe alguma informação histórica dos restaurantes, apesar de não ter sido obtida através de dados reais e não existir uma *database marketing* que permita um estudo diferente através de algoritmos preditores. O estudo vai ser assim feito a partir das características de cada restaurante, sendo possível através dos diferentes modelos atribuir um *score* final, que procura espelhar a qualidade do restaurante.

Num estudo futuro os algoritmos preditores podiam ser usados para prever a qualidade de cada cliente. Calculando em função dos seus atributos, por exemplo, a probabilidade de um cliente ir ao restaurante sabendo a sua profissão, idade, sexo, última vez que foi a um restaurante, entre outros.

Capítulo 3

Plataforma de Simulação

3.1. Vantagens de existir uma Plataforma de Simulação

Como podemos constatar, o Lead Management é um ciclo onde são congregadas várias etapas interligadas entre si. Apesar de algumas etapas não serem concretizáveis num curto espaço de tempo, alguns dos métodos usados nas mesmas têm de ser executados de uma forma rápida e precisa. A necessidade de existir uma plataforma de simulação que execute todas estas tarefas de uma forma agilizada é imprescindível.

Neste sentido, este trabalho assentou no desenvolvimento de uma plataforma de simulação onde pudessem ser estudados algumas das etapas e componentes do Lead Management, nomeadamente os modelos de *scoring*. Outra das vantagens do desenvolvimento de uma plataforma deste género é tornar possível estudar vários cenários diferentes e modelos de *scoring* criados em consonância com cada modelo.

O estudo partiu do pressuposto de que a fase de geração de *leads* já foi efectuada, apesar de este ser um trabalho contínuo.

3.2 Ambiente de desenvolvimento

A plataforma de simulação foi desenvolvida em Matlab usando um interface GUI. Foi escolhido o *software* Matlab por este se basear em cálculo numérico e possuir já muitas das funções necessárias para o trabalho. Outra das razões para a escolha recair sobre o Matlab é a maior facilidade com que se corrige questões de alguma complexidade, ao invés destas mesmas mudanças serem mais complexas e demorosas em outras linguagens. Apesar das vantagens, possui, como se sabe, a desvantagem de ser uma linguagem interpretada e, por isso, ser mais lenta do que uma linguagem compilada. A não existência de uma *toolbox* para a base de dados SQL levou ao uso de ficheiros

3.3 Arquitectura da Plataforma de Simulação

A arquitectura divide-se em dois módulos principais:

- Plataforma: responsável pela importação dos clientes e restaurantes, fazendo a gestão destes ao longo de todo o processo. Guarda em tabelas a informação de todos os resultados. Encarregado de toda a componente gráfica da interface.
- Simulador: tem com papel gerar os dados dos clientes e restaurantes. Lança o cliente na plataforma e simula a conversão e o feedback.

3.4 Funcionamento da Plataforma de Simulação

Para as próximas secções, é necessário ter presente alguns dos conceitos que vamos dar conhecimento aqui, antes de estes serem detalhados em profundidade mais à frente no trabalho. Como se optou por desenvolver a plataforma de simulação em inglês, são usados alguns termos em inglês.

- **Conversion (Conversão)** – Chamamos conversão ao fenómeno do cliente ir efectivamente ao restaurante, neste caso o restaurante converte a lead. Ao acontecimento contrário, isto é, quando o cliente não vai a um restaurante seleccionado, dizemos que não existiu conversão
- **Conversion Rate (Taxa de Conversão)** – É a taxa de conversão de cada restaurante dado um cliente. É o resultado da divisão entre o número de conversões pelo número de selecções.
- **Review Score** – É o valor dado pelo cliente no fim de ir ao restaurante, tendo em conta a sua satisfação perante o serviço. O valores possíveis a serem dados são entre 1 e 5.
- **Profile (Perfil)** – Cada cliente tem um perfil entre três possibilidades: demanding (exigente), neutral (neutro), friendly (simpático).
- **Customer (Clientes)** – Os *Customers* são os clientes que estão interessados em ir a um restaurante.
- **Provider (Prestadores de Serviço)**- Os *Providers* neste estudo são os restaurantes que estão registados na plataforma de simulação.

- **Plataform (Plataforma)** – A *Plataform* é o *software* que suporta os clientes, os restaurantes e os algoritmos de scoring e bonificação usados pelo simulador.
- **Simulador (Simulador)** – No *Simulador* geram-se os clientes e restaurantes. É onde se encontram os *modelos de scoring* e os algoritmos de bonificação das taxas de conversão.

De forma a ilustrar o comportamento geral do sistema, podemos observar na figura 2, o diagrama de arquitectura da Plataforma de Simulação.

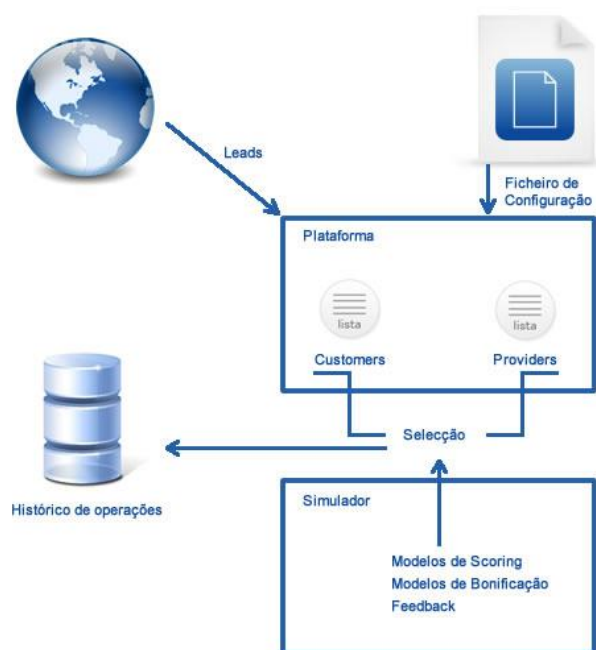


Fig. 2 – Diagrama de Arquitectura

A Plataforma de Simulação assenta em quatro actores principais: *Customers*, que representam os potenciais clientes; *Providers*, que são os prestadores de serviço, neste caso os restaurantes; Simulador, que simula todas as acções necessárias para o desenrolar do processo; Plataforma, que suporta os dados e funções dos anteriores agentes. Na figura 3 e 4 podemos verificar as operações que estão destinadas a cada actor.

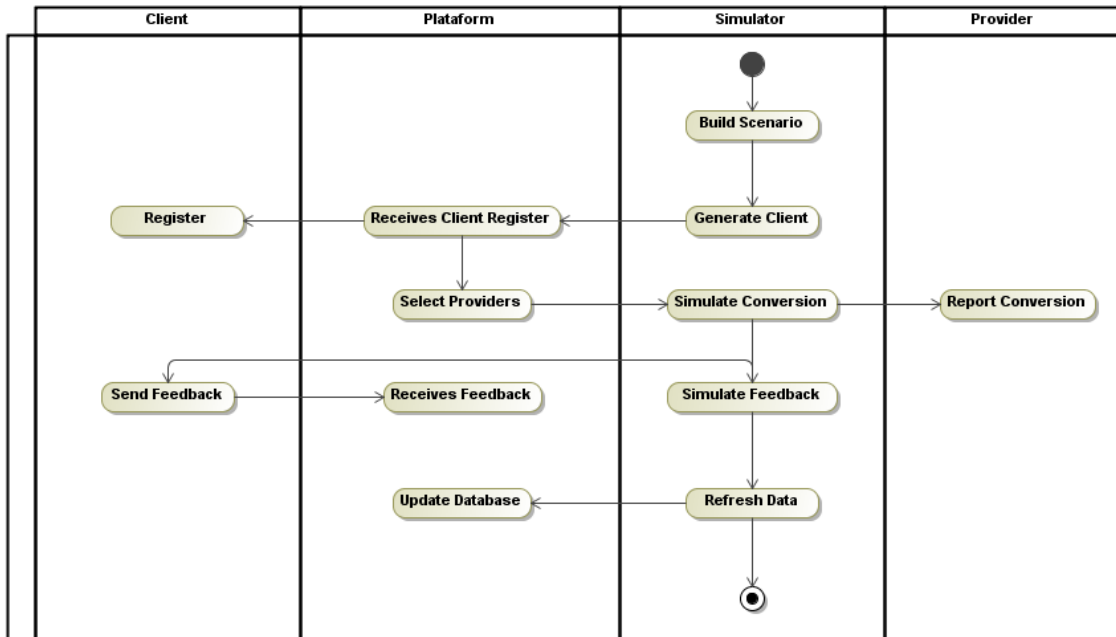


Fig. 3 – Diagrama de actividades

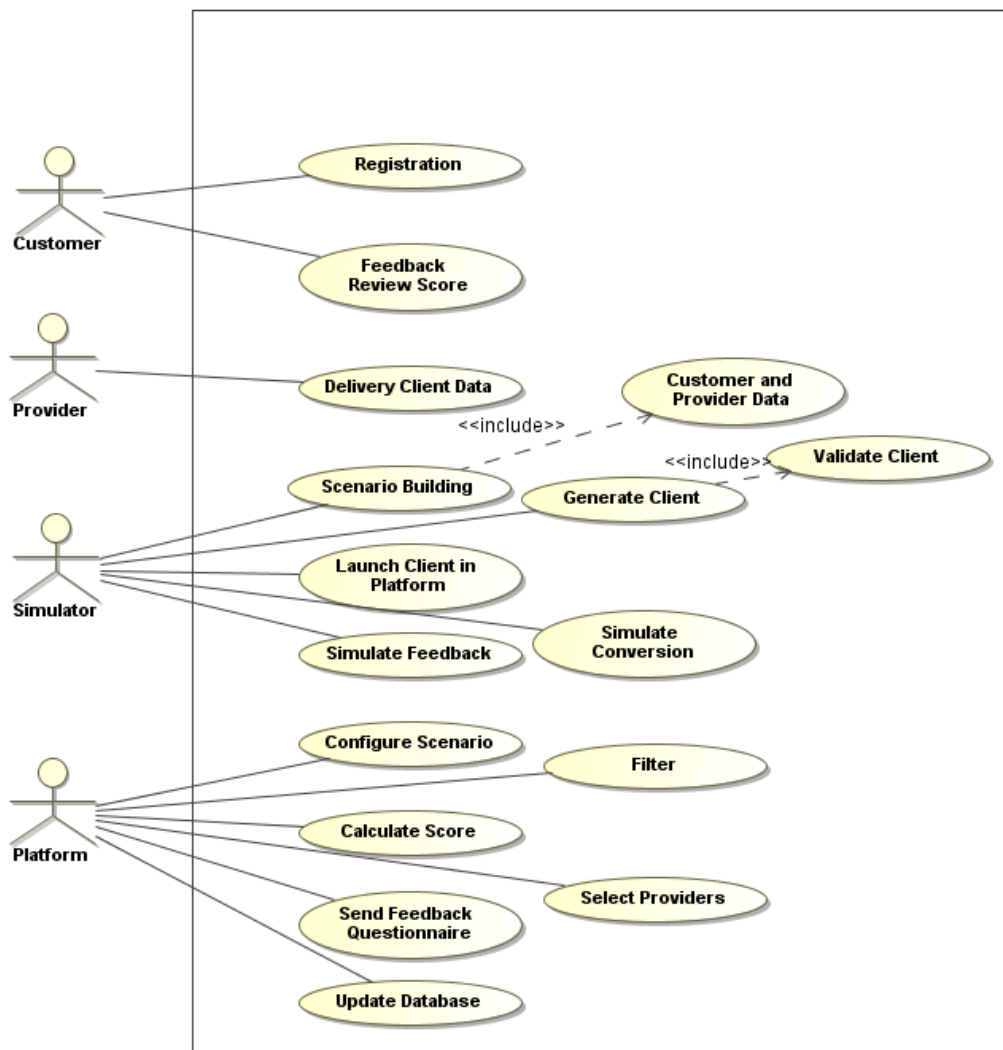


Fig. 4 – Diagrama de Casos de Uso de cada um dos agentes

Após a identificação dos principais actores foi necessário gerá-los. Os *Customers* (restaurantes) e *Providers* (clientes) são representados por um conjunto de atributos que simulam o mais fielmente possível pessoas e restaurantes, respectivamente. Os principais atributos escolhidos para gerar um *Customer* foram: *name*, *address*, e *profile*. Os *providers* têm como características principais: *address*, *conversion rate*, *review score* e *price*. Mais tarde na secção: Construção do Cenário, serão descritos todos os atributos bem como a forma como foram gerados.

Posteriormente à reprodução dos restaurantes e clientes foi necessário construir uma plataforma com uma interface gráfica de forma a ser possível carregar os dados e geri-los. A plataforma foi dividida em 4 áreas: *Customers* (1), *Providers* (2), Simulação (3) e Monotorização (4), tal como é ilustrado na figura 5. Na primeira área é onde se faz a importação dos *Customers* e se indica quantos destes vão ser lançados na plataforma. Aqui também se pode ter acesso a toda a informação de cada um dos *Customers*. No espaço destinado aos *Providers* faz-se o *load* dos restaurantes existentes e pode-se ver as características de cada um separadamente. Na secção de monotorização é mostrada uma tabela com todos os restaurantes, sendo possível ordenar automaticamente por cada um dos seus atributos. São ainda mostrados alguns dos parâmetros de configuração do simulador. No último espaço são identificados quais os modelos de *scoring* e bonificação a ser executados na simulação.

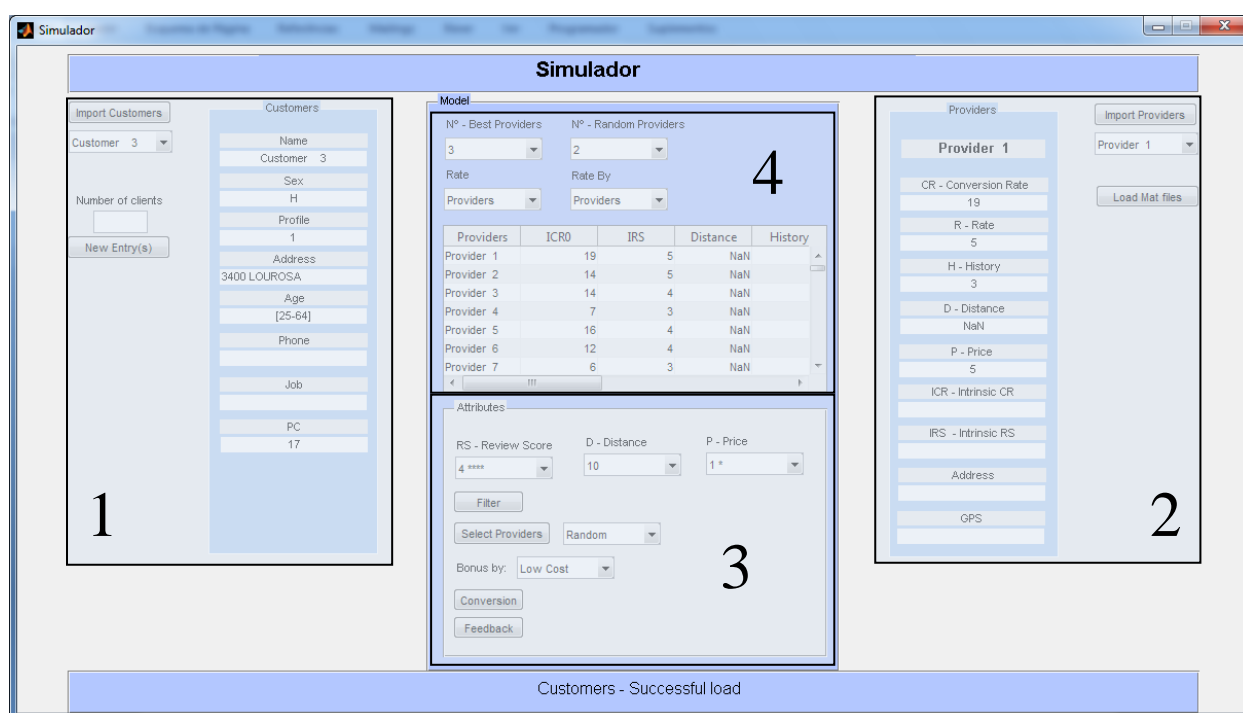


Fig. 5 – Imagem da Framework a destacar as principais áreas

A partir deste ponto a plataforma encontra-se preparada para suportar as funcionalidades do simulador e calcular todos os algoritmos necessários. O modo de funcionamento é cíclico, um ou mais clientes são lançados na plataforma e através dos seus atributos são seleccionados os restaurantes que se esperam que satisfaçam melhor o seu desejo através dos modelos de *scoring*. Depois de seleccionados pela plataforma, o simulador determina a existência ou não da conversão. Se existir uma conversão há uma probabilidade do simulador calcular um feedback dado pelo cliente. O processo termina sempre com actualização dos dados nos ficheiros de suporte que são a base do estudo, uma vez que é lá que se encontram os resultados das conversões, selecções e do feedback.

3.4.1 Customers

Como anteriormente referido, o passo de geração de *leads* é assumido como feito neste estudo. O número de clientes gerados para este estudo foram 50 mil, um número bastante considerável para a oferta existente (500 restaurantes).

Apesar do processo da geração de *leads* não ser feito, é demonstrado pela figura 6. O cliente ao se registar num *site* (ex: *micosite*) através de um formulário, insere os dados pessoais, contacto telefónico e os requisitos pretendidos para os restaurantes. Estes requisitos vão ser usados posteriormente na filtragem efectuada pela plataforma, onde serão eliminados todos os restaurantes que não satisfaçam os mesmos. A informação é recolhida e guardada numa base de dados.

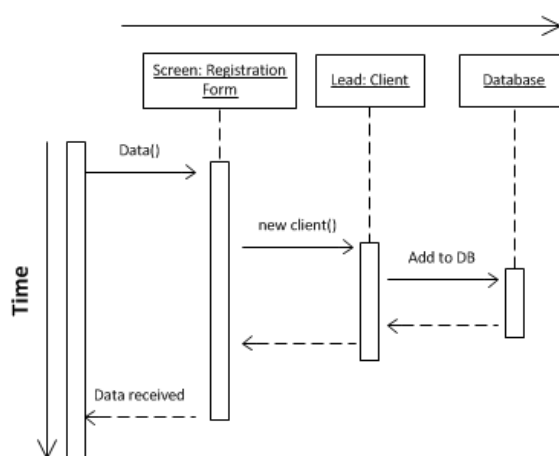


Fig. 6 – Diagrama de Interação – Registo no *site*

Os clientes, além do registo, recebem um questionário opcional de satisfação caso realmente tenham ido a um dos restaurantes. O Review Score (RS) é dado tendo em conta o *Profile* (perfil) do cliente e o RS obtido até à data pelo restaurante. Estas actividades são cíclicas e são repetidas para cada cliente como podemos verificar pela figura 7.

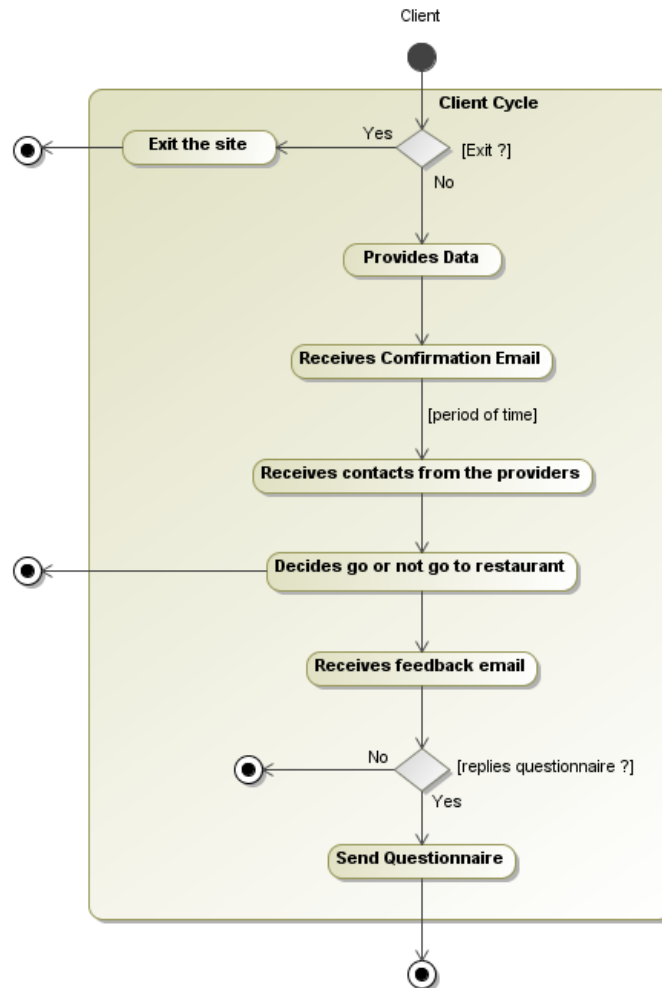


Fig. 7 – Diagrama de actividade do cliente

3.4.2 Providers

Os *providers* apenas têm como missão contactar os clientes que são indicados pela plataforma de simulação e dar informação à plataforma caso o cliente consume a ida ao restaurante. Esta informação é necessária para que sistemas que se baseiam em ser intermediários entre os clientes e os prestadores de serviço possam realmente ser viáveis. Uma vez que a empresa que detém a plataforma de geração e gestão de *leads*, obtém lucro apenas quando um cliente proveniente da sua captação vai a um restaurante, inscrito na sua plataforma.

Nesta troca de informação pode ocorrer um fenómeno de fraude, isso acontece quando o cliente vai ao restaurante mas este não o reporta ao gestor da plataforma. Apesar do restaurante poder lucrar como este facto, a plataforma penaliza-o automaticamente na percentagem de conversão do mesmo.

O fenómeno de fraude não vai ser estudado neste trabalho, porque fica fora do âmbito da dissertação. Fica apenas a referência de poderem existir casos desta natureza.

3.4.3 Plataforma

Como podemos observar na figura 8, a Plataforma tem como papel gerir todo o processo. Para isso, inicialmente existe uma configuração do cenário e a definição de algumas variáveis usadas pelo simulador. É gerido ainda, a recepção e encaminhamento das leads presentes na base de dados.

O primeiro passo para a selecção final dos restaurantes é a filtragem, onde são excluídos todos os restaurantes que não satisfazem o pedido feito pelo cliente no acto de registo. No registo, o cliente relativamente ao *RS*, escolhe o valor mínimo que pretende, entre 1 e 5. O mesmo acontece para a *Distance* mas no sentido oposto, aqui o cliente define a distância máxima entre si e os restaurantes, podendo ser: 10, 20, 30, 40 e 50 km no máximo. Por fim, relativamente ao *Price*, o cliente opta por um intervalo de gamas de preço: [1,2], [2,3], [3,4], [4,5].

Posteriormente a plataforma calcula o *score* dos restaurantes através dos modelos estudados e definidos no capítulo 6. Uma vez o score calculado, a plataforma selecciona os melhores restaurantes e no fim de feita a selecção a plataforma informa os restaurantes seleccionados, que por sua vez contactam os clientes. Por fim decorre uma actualização dos dados.

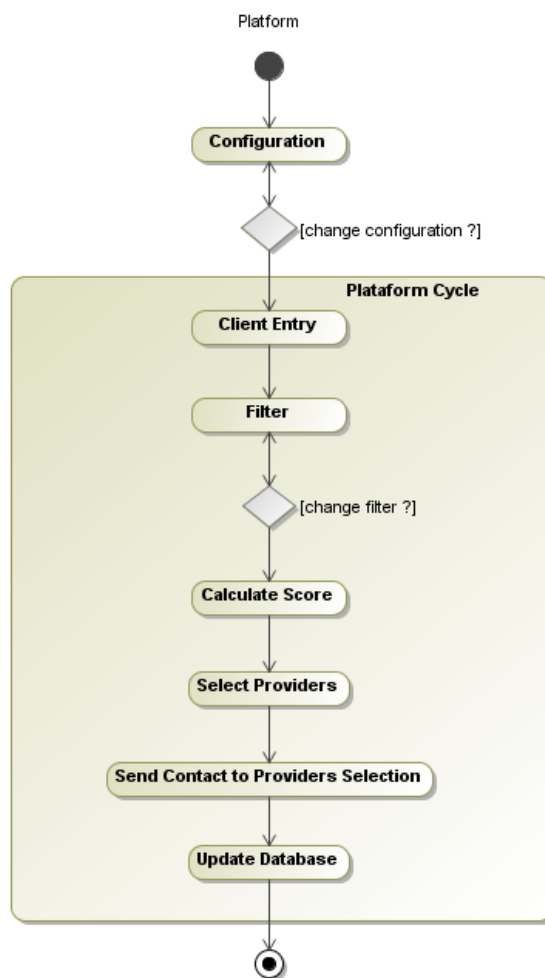


Fig. 8 – Diagrama de actividades da Plataforma

3.4.4 Simulador

É no simulador que se encontram as principais funções da plataforma de simulação (ver figura 9); é onde são gerados os dados dos clientes e restaurantes. É ainda, responsável também pela escolha de quais os clientes que dão entrada na plataforma. São escolhidos um conjunto de clientes entre os 50 mil existentes, sendo de referir que um cliente não pode ser seleccionado duas vezes numa simulação.

É aqui que se encontram os modelos de *scoring* que indicarão à plataforma quais os restaurantes a seleccionar.

No simulador são ainda calculadas as bonificações para os restaurantes seleccionados e determina-se se o cliente efectivamente vai ao restaurante. Esta decisão é simulada pelo algoritmo de conversão (ver abaixo), sendo que a probabilidade de o cliente converter é a soma das taxas de conversão dos restaurantes seleccionados, dado:

$$PC = \sum_{i=1}^5 CR(i) \quad (1)$$

Logo, a probabilidade de não converter:

$$PNC = 1 - PC \quad (2)$$

- **Algoritmo de conversão:**

- ✓ Calcula-se a PC e a PNC de (1) e (2):
- ✓ Normalizando entre 0 e 1

Neste momento, as probabilidades de conversão de cada restaurante e a probabilidade de não existir conversão podem ser esquematizadas desta forma:

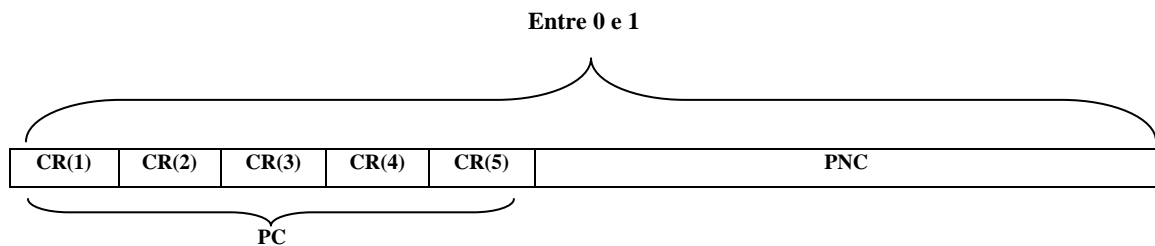


Fig. 9 – Esquema que representa as CR dos restaurantes seleccionados a e probabilidade de não existir conversão

- ✓ Com o auxílio da função rand() do matlab é gerado um valor entre 0 e 1, o resultado dirá, se existe ou não conversão, e em caso afirmativo qual é o restaurante.

Por fim, o simulador é ainda incumbido de determinar o feedback através do algoritmo de simulação (em anexo). O feedback é calculado através do *Profile* do cliente e do RS do restaurante.

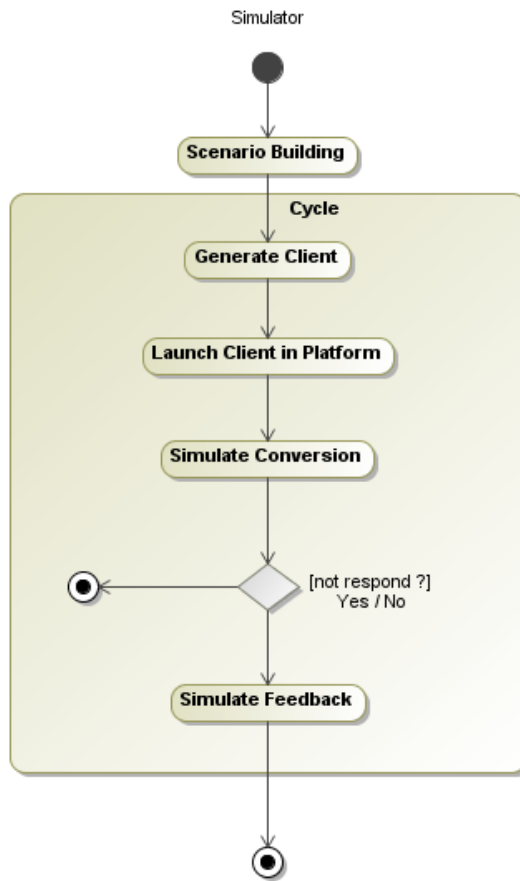


Fig. 10 – Diagrama de actividade do Simulador

Capítulo 4

Cenário

Após tomarmos conhecimento do funcionamento geral da Framework, neste capítulo vamos detalhar como foram gerados os restaurantes e os clientes bem como quais os seus atributos e o relevo de cada um para o estudo.

Os atributos gerados procuram replicar alguns dos comportamentos humanos, apesar de ser praticamente impossível, por estes estarem sujeitos a inúmeros factores que não podem ser medidos. Devido a estes factores, este trabalho assenta no estudo em concreto deste cenário. O cenário criado para este estudo é apenas uma possibilidade de análise para o problema. A construção do simulador teve esse mesmo intuito de se poderem testar diferentes cenários partindo de pressupostos diferentes. Deste facto advém a possibilidade de para cada cenário poderem existir resultados diferentes e daí se poderem extrair conclusões diferentes. Apesar desta aparente contrariedade perante cada cenário, as conclusões podem ser ambas válidas por se partir de presunções e objectivos diferentes.

Este cenário em concreto foi criado a partir de valores que se consideraram como os mais razoáveis e prováveis, a partir de alguns estudos existentes, pelo senso comum e por distribuições geográficas do país. É também de salientar que os valores das variáveis são independentes entre si, não foram criadas quaisquer dependências entre variáveis de forma a condicionar os valores finais. A independência das variáveis foi comprovada através do método dos mínimos quadrados no matlab, onde o grau de dependência verificado entre as principais variáveis não ultrapassou o 1%, grau de significância de 99%. Este aspecto é apenas uma característica do cenário que pode ser alterado num estudo com um diferente cenário.

As tabelas seguintes são de extrema importância, uma vez que expressam e descrevem todas as variáveis presentes no cenário. As variáveis são identificadas, assinaladas a forma como foram criadas: computacionalmente através de um algoritmo usado no processo de simulação, através de distribuições probabilísticas na criação do cenário ou se as variáveis são de configuração, introduzidas antes do início da simulação. A tabela 1 e 2 indicam ainda o tipo de variável: int, float, double e *three state*, a descrição estatística, os intervalos dos dados, uma descrição sucinta e os valores por defeito. Na tabela 2, podemos verificar a existência de uma coluna denominada *Filtered*, que indica quais os atributos que são filtrados pela plataforma.

Tabela 1 - Variáveis relativas ao Simulador

	Variable	Source of Variable	Variable Type	Statistic Description	Type	Interval [min, max]	Description	Default Values		
								P1	P2	P3
Simulator	N - random providers	Config	int	Not applicable	Discrete	-	Number of providers selected by the simulator, randomly	2	-	-
Simulator	M – best providers by Final Score	Config	int	Not applicable	Discrete	-	Selected number of providers, by Average Final Score	3	-	-
Simulator	Final Score	Computed	float	Not applicable	Continuous	-	Weighted score attributes of providers: CR, RS, Distance and Price	-	-	-
Simulator	Client Profile	Config	int	Probability Distribution cp P(cp) -1 30% Demanding 0 20% Neutral 1 50% Friendly	Discrete	-	Customer Profile: friendly, neutral or demanding			
Simulator	INC – Intrinsic Number of Conversions	Config	int	Normal distribution $X \sim N(\mu, \sigma^2)$	Discrete	-	Number of Conversions for each restaurant	100	30	-
Simulator	ICR – Intrinsic Conversion Rate	Config	float	Normal distribution $X \sim N(\mu, \sigma^2)$	Continuous	-	Intrinsic Conversion Rate of the restaurant for each client	8	2.4	-
Simulator	IPRS – Review Score	Config	float	Normal distribution $X \sim N(\mu, \sigma^2)$	Continuous	[1,5]	Score given by the customer after conversion	3.5	1	-
Simulator	PPrice	Config	int	Qui-squared distribution $X \sim \chi_P^2(\lambda)$	Discrete	[1,5]	Restaurant Price	0.2	0.4	-

	Variable	Source of variable	Variable Type	Description	Filtered
Client	Has converted	Computed	three State	The customer goes to the restaurant or not	No
	Has reviewed	Computed	three State	The client replies to the sent questionnaire	No
	ReviewScore	Computed	int	-	No
	Name	Provided	string	-	No
	Address	Provided	string	-	No
	Job	Provided	string	-	No
	Birthday	Provided	string	-	No
	CProfile	Generated	int	-	No
	Sex	Generated	string	-	No
	Email	Provided	string	-	No
	Phone	Provided	string	-	No
Providers	Name	Provided	string	-	No
	Address	Provided	string	-	Yes
	PCR – Provider Conversion Rate	Computed	float	-	No
	PICR –Provider Intrinsic Conversion Rate	Generated	float	-	No
	NC – Number of conversions	Generated	int	-	No
	PRS – Provider Review Score	Computed	float	-	Yes
	PIRS –Provider Intrinsic Review Score	Generated	float	-	No
	PPrice	Generated	int	-	Yes
	GPS	Generated	float [2]	-	No

Tabela 2 – Variáveis relativas ao Clientes e Restaurantes

4.1 Construção do Cenário

Depois de identificar os actores da plataforma de simulação, saber as tarefas a que cada um está sujeito ao longo processo e identificar as variáveis usadas no cenário, vamos agora detalhar os atributos expressos nas tabelas dos clientes, restaurantes e simulador.

Na maioria dos atributos foram usadas distribuições de probabilidade normal. A escolha recaiu sobre esta distribuição por ser fundamental na teoria estatística em diversas áreas e ser facilmente definida, sendo apenas necessário dois parâmetros: a média μ e o desvio padrão σ . A um desvio padrão da média estão inseridos cerca de 68% da amostra total e a cerca de dois desvios padrão está cerca de 95% da amostra, ou seja, os valores extremos verificam-se poucas vezes.

Como a função que define a distribuição normal é assintótica, foram gerados alguns valores negativos. Uma vez que não podem existir valores negativos, estes foram desprezados, sem com isto alterar significativamente a média de todos os valores.

▪ **Simulador**

✓ **Random Providers e Best Providers**

As variáveis usadas no módulo do simulador, os Random Providers (N) representam o número de restaurantes escolhidos pelo modelo de selecção de modo aleatório, os Best Providers (M), indicam o número de restaurantes a serem seleccionados pelo modelo, mas aqui são seleccionados os M melhores restaurantes.

▪ **Clientes**

Os atributos escolhidos que caracterizam os clientes são o *Profile* e *Feedback Probability*, além dos atributos dos dados pessoais que não têm relevância para este estudo.

✓ **Profile**

O *Profile* do cliente tem três estados diferentes: *demanding*, *neutral* e *friendly*. O perfil do cliente é importante para o valor dado no RS, um cliente exigente por norma dá uma nota mais baixa do que o nota que o restaurante tem no momento, já um cliente neutral dá uma nota

próxima do valor do RS intrínseco do restaurante. No caso de o cliente ser simpático por norma dá uma nota superior ao RS intrínseco do restaurante.

✓ **Feedback Probability**

O *Feedback Probability* define a probabilidade do cliente responder ao questionário de satisfação, esta variável tem um valor fixo de 15%.

▪ **Restaurantes**

O número de restaurantes gerados foi 500, ou seja, dá uma média de um restaurante para cada 100 pessoas.

Os restaurantes são definidos por quatro variáveis distintas, *Intrinsic Conversion Rate* (ICR), *Intrinsic Review Score* (IRS), *Distance* e *Price*.

✓ **Intrinsic Conversion Rate**

A ICR representa a taxa intrínseca de conversão de cada restaurante, para gerar os valores foi usada uma distribuição normal: $N(8,2.4)$.

✓ **Review Score**

O *Review Score* do restaurante é a média de todos os scores dados pelos clientes após a concretização.

✓ **Intrinsic Review Score**

O IRS está compreendido numa escala de 1 a 5. Os dados são gerados por uma distribuição normal: $N(3,0.7)$, não ocorrem quaisquer valores negativos.

✓ **Distance**

A *Distance* é a distância em Km entre o endereço do cliente e do restaurante. É calculada através de cada par de códigos postais, pela determinação das coordenadas geográficas.

Seja P_1 e P_2 os locais expressos em coordenadas geográficas $P_1 = (\phi_1, \lambda_1)$ e $P_2 = (\phi_2, \lambda_2)$. Pela trigonometria esférica o arco de uma circunferência S entre os pontos P_1 e P_2 é expresso pela seguinte equação (Frank Ivis, 2006):

$$\cos(S) = \sin(\phi_2) \sin(\phi_1) + \cos(\phi_2) \cos(\phi_1) \cos(\Delta\lambda) \quad (1)$$

Após o cálculo do arco da circunferência, é possível determinar o seu comprimento. Considerando o raio médio terrestre (R) de 6378 km, a distância entre P_1 e P_2 é expressa por (Frank Ivis, 2006):

$$dist_{P_1P_2} = \alpha \cos(S) * R = \alpha \cos(S) * 6378 \quad (2)$$

Apesar dos raios terrestres não serem constantes ao longo de diferentes latitudes, para pequenas distâncias, como é o caso, o erro gerado não é significativo, por isso a opção de usar o raio terrestre médio. De referir que as coordenadas geométricas foram encontradas pela API do Google Maps usando um script em php disponibilizado pela Google, o script foi alterado na conexão à base de dados bem como no tempo de execução do mesmo.

✓ **Price**

O *Price* dos restaurantes é de 1 a 5, tendo em conta os preços praticados pelos restaurantes. Neste caso os dados são gerados por uma distribuição normal: $N(3,0.6)$, também não ocorrem quaisquer valores negativos. Apesar de a distribuição ser definida por uma função contínua e conseqüentemente não gerar valores discretos, foram arredondados de forma a existirem apenas valores discretos

✓ **Intrinsic Number of Conversions (INC)**

O número de conversões indica quantos clientes cada restaurante já captou. Os valores são aleatoriamente gerados entre 30 e 100 clientes.

4.1.1 Inserção esporádica de novos restaurantes

O cenário contempla ainda o estudo da inserção de novos restaurantes na plataforma. É mais um aspecto que permite estudar a competência e eficácia dos modelos de *scoring*. O

objectivo é diferenciar estes restaurantes da melhor forma sabendo que à partida não existe informação sobre eles.

Neste estudo, por cada entrada de um novo cliente, a plataforma calcula através da probabilidade de $1/500$ se existe ou não um novo restaurante. Em média por cada 500 novos clientes é criado um novo restaurante.

Uma vez que sobre estes restaurantes não existe qualquer informação, inicialmente estes restaurantes terão nos valores dos seus atributos, nomeadamente na CR e RS, a média de todos os restaurantes já existentes. Estes valores vão perdurar até que sejam calculados os valores efectivos, isso acontece, quando o novo restaurante converter vinte vezes. Foi escolhido este valor, uma vez que para um restaurante obter este número de conversões são necessários no mínimo cerca de 20 mil clientes, dando assim tempo de a taxa real calculada, reflectir o mais fielmente possível o “valor” do restaurante.

Capítulo 5

Modelos de *scoring* e bonificações das taxas de conversão

5.1 Modelos de *scoring*

Como anteriormente se viu, os algoritmos de *scoring* baseiam-se no historial dos clientes ou empresas de modo a extrair informações privilegiadas. Neste caso, os algoritmos de *scoring* vão identificar os melhores restaurantes por diversos critérios. Os resultados do número de conversões indicará qual ou quais os modelos que melhor servem os interesses de ambos os lados, clientes e restaurantes.

Os modelos criados são centrados do lado da oferta, ou seja, dos restaurantes, sendo o objectivo seleccionar os melhores restaurantes de forma a aumentar as idas dos clientes aos mesmos. A escassa informação dos clientes será apenas utilizada nos algoritmos de bonificação na CR. Para se evidenciar as diferenças entre modelos, foram criados modelos simples e modelos mais complexos. É considerado um algoritmo simples aquele que apenas considera um atributo para a ordenação final; um modelo complexo entra com diversos atributos.

5.1.1 Modelos Simples

Os modelos simples foram usados para se perceber de facto o efeito de cada variável no número de conversões global e até que ponto os modelos que integram todos os atributos são mais eficazes. Em caso de existirem restaurantes em que o valor do atributo em análise é igual, o desempate é feito sempre através da menor distância. Desta forma foram criados os seguintes modelos:

✓ **Modelo: *Random***

É um modelo em que todos os restaurantes são seleccionados de forma aleatória após a filtragem, aqui os atributos entram apenas para filtrar os restaurantes que satisfazem os requisitos dos clientes. A variável N que define o número de restaurantes seleccionados de forma aleatória toma o valor 5.

✓ **Modelo: *Conversion Rate***

Assim como o nome indica, este modelo selecciona os restaurantes com melhor *Conversion Rate*. Espera-se que com este modelo um maior número de conversões no final. À excepção do modelo anterior, aqui a variável M toma o valor 3 e a variável N o valor 2.

✓ **Modelo: *Review Score***

Assim, tal como no modelo anterior, o modelo *Review Score* selecciona os restaurantes com os melhores *Review Scores*.

✓ **Modelo: *Distance***

Como anteriormente vimos, a cada cliente e a cada restaurante foi atribuída uma localização através do par código postal-freguesia. Através da localização dos clientes e restaurantes são calculadas as distâncias entre si. O modelo *Distance* selecciona os M primeiros restaurantes, ou seja, os restaurantes que estão mais perto do cliente.

✓ **Modelo: *Price***

O modelo *Price* selecciona os restaurantes da gama de preços mais elevada de entre os filtrados.

5.1.1 Modelos Complexos

Nos modelos complexos são esperadas taxas de conversão tendencialmente superiores à probabilidade de conversão (PC) média, por se identificarem os melhores restaurantes em mais do que uma característica simultaneamente.

✓ Modelo: *Mean*

No modelo *Mean* é calculado um score final de 0 a 1 tendo em conta a PC, RS, Distance e Price. Cada atributo tem um peso (p) de 25%. Todas as variáveis têm o mesmo peso para o *score* final, sendo que uma característica menos boa por parte do restaurante pode ser camuflada pelos restantes atributos. As diferenças entre os restaurantes não são muitas evidenciadas.

$$Score = p * CR(i) + p * RS(i) + p * Distance(i) + p * Price(i) , i=1,...,M \quad (3)$$

✓ Modelo: *Top Average*

O modelo *Top Average* faz uma selecção mais apertada dos restaurantes, apenas os restaurantes que registarem valores mais altos nos seus atributos terão alguma cotação. Este modelo não é expresso por uma fórmula como o anterior, foi efectuado através de ciclos *for*. O algoritmo em pseudo-código representa os diversos ciclos desenvolvidos no cálculo do score.

```
Para CR  
Se  $CR(i) > Média(ICR) + Média(ICR)/2$ ;  
   $CR\_Score = 0.25$ ;  
Caso contrário  
   $CR\_Score = 0.0$ ;  
  
Para RS e Price (exemplo RS)  
Se  $RS(i) > 4$ ;  
   $RS\_Score = 0.25$ ;  
Caso contrário  
   $RS\_Score = 0$ ;  
  
Para Distance  
Se  $Distance < 10$ ;  
   $Distance\_Score = 0.25$ ;  
Caso contrário  
   $Distance\_Score = 0.0$   
  
Score =  $CR\_Score + RS\_Score + Price\_Score + Distance\_Score$ 
```

Se a CR do restaurante for superior à soma da média da ICR de todos os restaurantes mais metade deste valor, tem um CR_Score de 0.25. Por exemplo, um restaurante com uma taxa de conversão de 13%, tem mais do que os 12% necessários para obter o *score*, porque a média do ICR, mais a média do ICR sobre dois é 12% ($8 + 8/2 = 12\%$).

No caso do RS e do Price, se estes forem superiores a 4, ou seja, igual a 5, RS_Score é igual a 0.25. Finalmente para a distância o modelo *Top Average*, atribui um $Distance_Score$ de 0.25 para distâncias inferiores a 0.25.

✓ Modelo: BCG - 1

Este modelo é fundamentado na matriz BCG¹, que é usada no marketing para se estudar o ciclo de vida de um produto. Neste modelo também vão existir duas dimensões: o número de conversões (NC) e taxa de conversão (CR). O NC pode ser baixo, médio ou alto e a CR pode ser baixa, média ou alta. Através destas duas dimensões e dos seus graus quantitativos foi construída a seguinte matriz.

Score		Número de Conversões		
		Baixo <=20	Médio >20 e <100	Alto >=100
Taxa de Conversão	Baixa <= 6	0.5	0.4	0.6
	Média >6 e < 10	0.6	0.5	0.8
	Alta >= 10	0.7	0.6	1

Tabela 3 – Matriz de pesos para cálculo do Score

Para cada cruzamento é dado um score ao restaurante. Como podemos observar existem algumas diferenças nos valores, porque são atribuídas importâncias diferentes a cada par de entradas. Considera-se que restaurantes com pouco número de conversões e com taxas altas de conversão são melhores que restaurantes com um número médio de conversões e com taxas de conversão igualmente altas. Ou seja, os restaurantes com número de conversão baixos têm mais potencial que restaurantes com um número médio de conversões para a mesma taxa de conversão. Neste modelo, o *score* final é quanto maior quanto maior a CR mas o NC não segue uma forma linear, como podemos observar na figura 11.

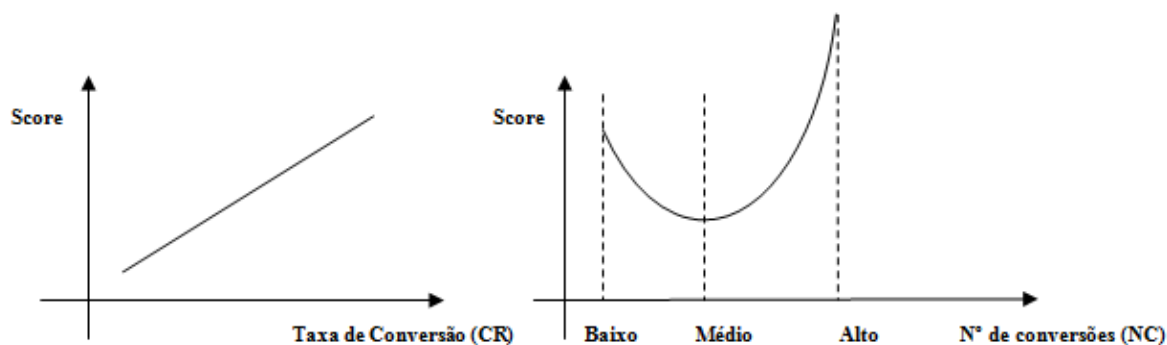


Fig. 11 – Gráficos que elucidam a forma como o Score evolui com o aumento da CR e do NC

¹ - Desenvolvida pela Boston Consulting Group

5.2 Bonificação da Taxa de Conversão

Uma vez que os algoritmos de *scoring* são calculados através dos atributos dos restaurantes, foram criados os modelos de bonificação de forma a simular o aumento de percentagem de concretizações, entre os restaurantes seleccionados, caso estes possuam as características desejadas pelo tipo de cliente.

Tipicamente diferencia-se a população em três tipos de classe consoante a sua situação económica: classe baixa de poucos rendimentos, classe média no qual estão a maioria das pessoas e classe alta em que se encontram uma reduzida percentagem de pessoas. Foi segundo esta perspectiva que foram criados os modelos de bonificação, *Low Cost*, *Quality vs Price* e *High Quality*.

5.2.1 Low Cost

Podemos aferir um restaurante *low cost*, por ter preços baixos e por se situar na proximidade do cliente, de forma que a deslocação seja pequena. Aos restaurantes a menos de 10 km do cliente com um Price igual a 1 é atribuído um bónus de 2%. Caso o restaurante esteja a menos de 20 km e tenha um Price menor do que 3 tem um bónus de 1%

Bonificação		Distance	
		Pequena <=10	Média-Baixa <10 e <=20
Price	Baixo 1	2%	1%
	Médio - Baixo >1 e < 3	1%	1%

Tabela 4 – Matriz com duas entradas (*Distance* e *Price*) para calcular o valor da bonificação

5.2.2 Quality vs Price

Este modelo procura essencialmente os restaurantes com uma boa qualidade-preço para os clientes de classe média. Para aferir a qualidade usa-se o RS e o CR e para aferir o preço a *Distance* e o *Price*. Para o restaurante ter uma bonificação de 2% na CR, tem de se encontrar a menos de 10km do cliente e possuir um Price igual a 3, simultaneamente o RS e CR não pode ser

inferior a 3.5 e 8% respectivamente. O restaurante tem um bônus de 1% se tiver um *Price* entre o 2 e 4 , ficar a menos de 20 km, ter um *RS* superior a 3 e um *CR* superior a 8%.

Bonificação		CR			Distance
		Elevada >=12%	Média-Alta >8 e <12		
RS	>3.5	2%	1%	<=10	
	>3	1%	1%	>10 e <20	
	= 3		>2 e <4		
Price					

Tabela 5 – Matriz com quatro entradas (*CR*, *RS*, *Price* e *Distance*) para calcular o valor da bonificação

5.2.3 High Quality

Para responder aos clientes de classe mais alta, foi criado o modelo *High Quality*, que premeia os restaurantes com boas classificações na taxa de conversão e ter preços elevados. Para existir uma bonificação de 2% os restaurantes têm de ter uma *CR* superior a 12% e um *Price* de 5. Existe uma bonificação de 1% quando a *CR* é superior a 8% e o *Price* maior ou igual a 4.

Bonificação		CR	
		Elevada >=12%	Média-Alta >8 e <12
Price	Alto 5	2%	1%
	Médio - Alto 4	1%	1%

Tabela 6 - Matriz com duas entradas (*Price* e *CR*) para calcular o valor da bonificação

Capítulo 6

Testes e Resultados

6.1 – Testes

De modo a testar o comportamento dos modelos de *scoring*, sobre a evolução da taxa de conversão sobre os restaurantes foram criados três tipos de testes:

1. Tem como principal objectivo, saber qual ou quais os modelos em que existe um maior número de conversões.
2. O objectivo deste tipo de teste é analisar a evolução da taxa de conversão dos novos restaurantes no decorrer de uma simulação.
3. Estudar três restaurantes em particular, com diferentes taxas de conversão, examinando em detalhe a evolução das suas taxas de conversão ao longo de uma simulação.

▪ Tipo 1

São lançados 25 mil clientes na plataforma por cada combinação: Modelo de Scoring/Algoritmo de bonificação. Uma vez que existem 8 modelos e 3 algoritmos, existem 24 testes do tipo 1. Nestes testes não se estuda a inclusão de novos restaurantes.

▪ Tipo 2

Através dos resultados dos Testes do Tipo 1, são escolhidas os 2 modelos com maior número de conversões e o modelo *Random*, para analisar a inclusão de novos restaurantes. Nestes testes também são lançados 25 mil clientes.

▪ Tipo 3

São estudados três restaurantes ao longo das mesmas 3 simulações efectuadas nos testes anteriores, comparando estes resultados com os do Tipo 2, uma vez que são efectuados nas mesmas condições.

6.2 Resultados

Após se ter efectuado os diferentes tipos de testes, foram analisadas as tabelas onde são registados os resultados das simulações.

▪ Tipo 1

Os resultados dos testes do Tipo 1, podem ser observados na tabela 7. Para cada teste foi contabilizado o número de conversões (NC), número de selecções (NS), número de *Review Scores* (RS), taxa de conversão (CR), ganho em relação à taxa de conversão intrínseca (ICR), entre outros que podem ser vistos em anexo na Tabela 7.

Modelos de Scoring - Bonificação	CR	NC	NS	NRS	Média CR	Médis RS	Ganho em relação há ICR
Random - Low Cost	8,26%	10043	121530	1595	8,04	3,48	3%
Random - P vs Q	8,05%	9798	121687	1529	7,78	3,50	1%
Random - HQ	8,08%	9827	121628	1433	8,00	3,51	1%
RS - Low Cost	8,47%	10369	122384	1524	7,98	3,53	6%
RS - P vs Q	8,37%	10229	122216	1558	8,01	3,56	5%
RS - HQ	8,32%	10180	122284	1515	7,81	3,55	4%
CR - Low Cost	10,49%	12836	122311	1949	8,17	3,53	31%
CR - P vs Q	10,56%	12923	122320	1902	8,09	3,48	32%
CR - HQ	10,29%	12592	122337	1922	8,00	3,57	29%
Distance - Low Cost	8,37%	10237	122348	1510	8,06	3,44	5%
Distance - P vs Q	8,69%	10630	122275	1595	8,56	3,46	9%
Distance - HQ	8,40%	10271	122282	1556	8,23	3,52	5%
Price - Low Cost	8,34%	10204	122315	1520	8,05	3,49	4%
Price - P vs Q	8,86%	10826	122190	1607	8,66	3,49	11%
Price - HQ	8,56%	10472	122330	1527	8,19	3,48	7%
Mean - Low Cost	9,20%	11253	122270	1762	8,01	3,56	15%
Mean - P vs Q	9,84%	12028	122268	1790	8,52	3,48	23%
Mean - HQ	9,43%	11539	122310	1688	8,24	3,59	18%
Top Average - Low Cost	8,59%	10505	122247	1578	8,22	3,59	8%
Top Average - P vs Q	8,74%	10687	122275	1589	8,29	3,60	9%
Top Average - HQ	8,61%	10523	122261	1611	8,11	3,61	8%
BCG - 1 - Low Cost	9,74%	11923	122352	1837	8,27	3,53	22%
BCG - 1 - P vs Q	10,21%	12486	122264	1879	8,62	3,51	28%
BCG - 1 - HQ	9,92%	12131	122266	1821	8,12	3,53	24%

Tabela 7 – Quadro resumo dos resultados do teste do tipo 1

O modelo onde se verificaram menos conversões foi no modelo *Random*, onde as taxas são idênticas (~8%) à média intrínseca de conversão de todos os restaurantes (ICR). Este modelo foi criado com o propósito de servir de referência aos resultados obtidos nos outros modelos. Um resultado inferior ao modelo *Random* significa que o modelo em estudo não é eficiente.

Como podemos observar na Tabela 7, o modelo de *scoring* que produz um maior número de conversões é o CR, onde os restaurantes são ordenados pela sua taxa de conversão. Este resultado pode ser justificado, uma vez que, quanto maior a taxa de conversão, maior é a probabilidade de existir uma conversão.

O modelo BCG-1 foi o segundo modelo onde se registaram mais conversões, a taxa de conversão ultrapassou mesmo os 10%, quando este modelo teve como bonificação – *Price vs Quality*. É de salientar que, nos oito modelos estudados, seis obtiveram a sua maior taxa de conversão com a bonificação – *Price vs Quality* (ver figura 12 e tabela 8).

Pelos resultados obtidos ao nível das taxas de conversão, podemos concluir que é possível aumentar a taxa de conversão através dos modelos estudados, em relação à taxa intrínseca.

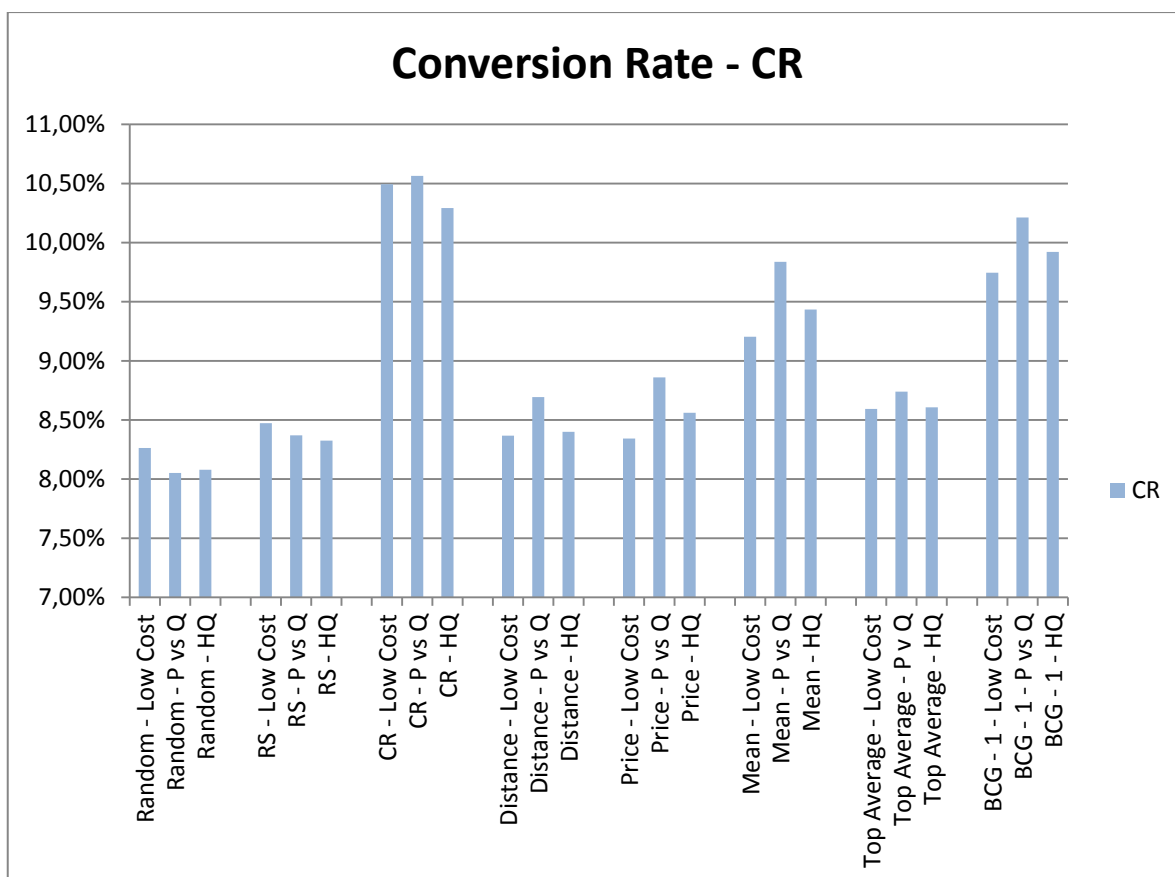


Fig. 12 – Gráfico com as taxas de conversão nas 24 combinações

	CR
Low-Cost	8,94%
Price vs Quality	9,17%
High Quality	8,95%

Tabela 8 – Taxa de conversão média por tipo de bonificação

Como seria de esperar, quanto maior é o número de conversões, maior é o número de *Review Scores* dados pelos clientes (ver anexo: Tabela X). Facto que se verifica uma vez que é no modelo CR onde se verifica o maior número de *Review Scores*.

Pelos valores médios dos *reviews scores*, podemos afirmar que estes são independentes do número de vezes que são dados (ver anexo: Tabela X), uma vez que este apenas depende do perfil (*profile*) dos clientes. O modelo onde a média real do *review score* foi mais elevada (3,61), foi no *Top Average*.

▪ Tipo 2

Nos testes do tipo 2 é estudado o comportamento dos novos restaurantes, quando estes são inseridos no decorrer da simulação. Após se saber os resultados dos primeiros testes, foram seleccionados os modelos CR e BCG-1, a somar ao modelo *Random* para servir de comparação. Todos eles tiveram como combinação de bonificação o *Price vs Quality*.

No caso do modelo *Random*, foram criados 56 novos restaurantes, no modelo CR surgiram 59 e por fim no modelo BCG-1 foram inseridos 53 restaurantes (ver tabela 9).

	Nº de restaurantes novos	Média CR	Máximo CR*	Mínimo CR*
Random - Price vs Quality	56	9,35%	10,44%	8,03%
CR - Price vs Quality	59	9,24%	11,03%	8,39%
BCG -1- Price vs Quality	53	8,02%	9,05%	6,85%

*Taxas relativas apenas aos novos restaurantes

Tabela 9 - Quadro resumo dos resultados dos novos restaurantes.

No modelo *Random*, a taxa de conversão média de todos os restaurantes teve a evolução verificada na figura 12. A média apenas é calculada quando todos os restaurantes são seleccionados pelo menos uma vez, pelo que de início não se registam quaisquer valores.

Inicialmente a taxa média de conversão apresenta muitas oscilações, até estabilizar a partir dos 10 mil clientes.

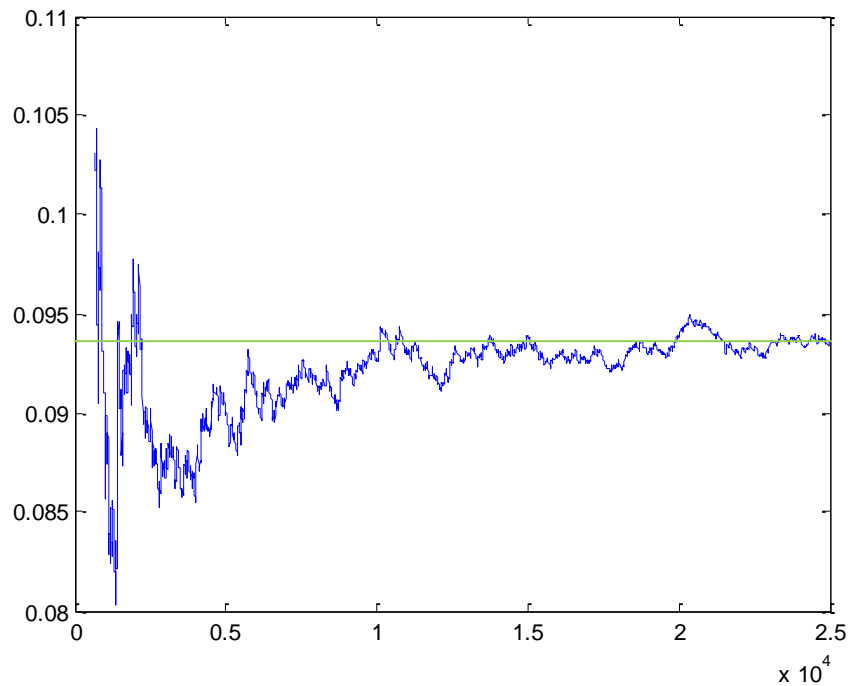


Fig. 13 – Gráfico da taxa de conversão média dos novos restaurantes pelo modelo *Random*

A taxa de conversão dos novos restaurantes no modelo CR teve um comportamento diferente do anterior. Primeiramente as taxas de conversão foram elevadas, superando os 10%, mas progressivamente foram baixando até cerca de 8,5%, até que no último terço da simulação a taxa subiu aproximadamente até aos 9,20%.

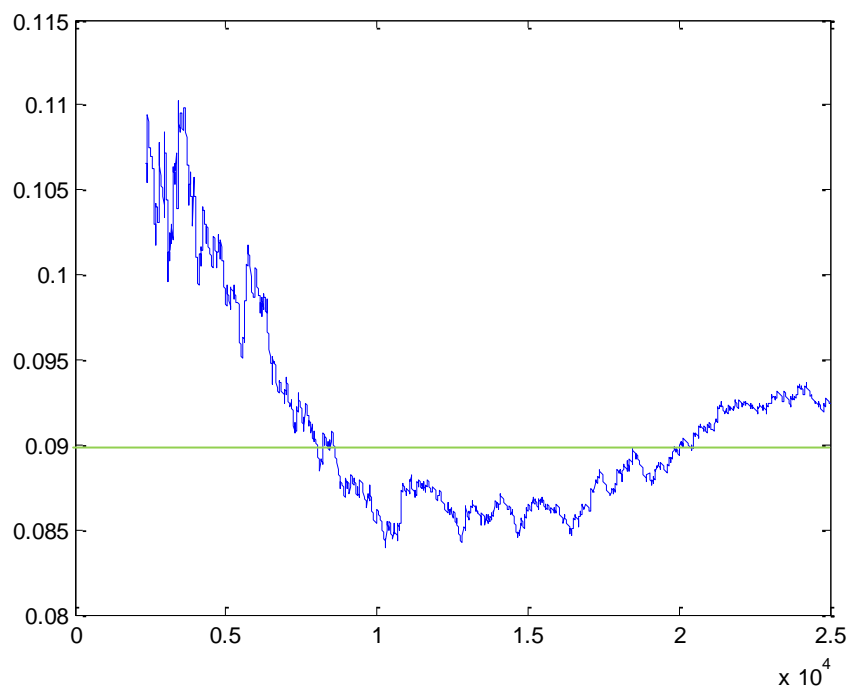


Fig. 14 - Gráfico da taxa de conversão média dos novos restaurantes pelo modelo CR

No último modelo BCG-1, a evolução da taxa de conversão seguiu o mesmo padrão do modelo CR, apesar de as taxas serem um pouco mais baixas. As taxas menos elevadas neste modelo podem ser explicadas por este ter em consideração o número de conversões dos restaurantes. Uma vez que os novos restaurantes não tem conversões, a probabilidade de virem a ser seleccionados diminui.

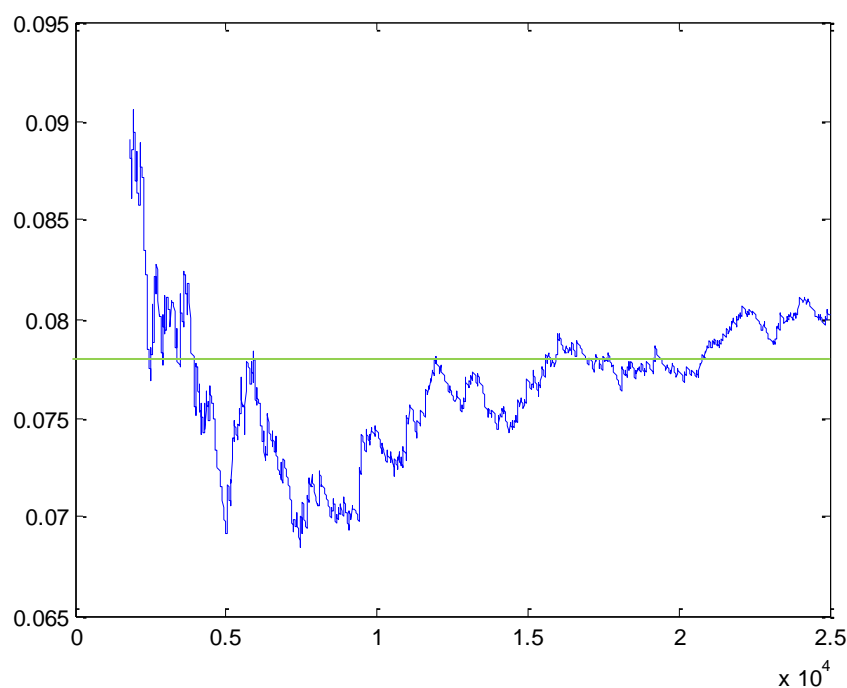


Fig. 15 - Gráfico da taxa de conversão média dos novos restaurantes pelo modelo BCG-1

Relativamente ao número de conversões e selecções apresentados pelos novos restaurantes, pode-se verificar pelas Tabelas 10, 11, 12 e 13, que apresentam valores significativamente menores do que nos restaurantes existentes. O modelo onde os valores do número de conversões e selecções são mais elevados é no modelo *Random*.

Como os novos restaurantes partem de uma ICR (taxa de conversão intrínseca) de 8%, a probabilidade destes serem seleccionados é elevada, em comparação aos restaurantes que já existem com taxas de conversões inferiores a 8%.

	Máximo (Nº conversões)	Mínimo (Nº de conversões)	Média (Nº de conversões)
Random - Price Quality	27	7	14,15
CR - Price Quality	11	0	6,47
BCG-1- Price Quality	12	1	5,5

Tabela 10 – Quadro resumo do nº de conversões dos novos restaurantes

	Máximo (Nº de vezes selecionado)	Mínimo (Nº de vezes selecionado)	Média (Nº de vezes selecionado)
Random - Price Quality	198	87	150
CR - Price Quality	103	41	68
BCG-1- Price Quality	107	43	68

Tabela 11 - Quadro resumo do nº de selecções dos novos restaurantes

	Máximo (Nº de conversões)	Mínimo (Nº de conversões)	Média (Nº de conversões)
Random - Price Quality	122	0	19,91
CR - Price Quality	390	0	26,07
BCG-1- Price Quality	237	0	25,36

Tabela 12 - Quadro resumo do nº de conversões dos 500 restaurantes existentes

	Máximo (Nº de vezes selecionado)	Mínimo (Nº de vezes selecionado)	Média (Nº de vezes selecionado)
Random - Price Quality	1092	2	227
CR - Price Quality	2659	0	237
BCG-1- Price Quality	3216	0	237

Tabela 13 - Quadro resumo do nº de selecções dos 500 restaurantes existentes

▪ Tipo 3

No último tipo de testes é estudado o comportamento da taxa de conversão em 3 restaurantes. Os restaurantes diferem na sua taxa de conversão (CR), o primeiro tem 8%, o segundo 3% e o último tem uma CR de 14%.

No primeiro restaurante (R1), o trajecto seguido pela taxa de conversão é diferente nos 3 modelos apesar de a taxa final ser idêntica nos três (ver figuras 16, 17 e 18). As CR dos 3 restaurantes subiram em média 2%.

No caso do segundo restaurante (R2), as taxas de conversão são muito voláteis, no caso do modelo *Random* e CR as taxas finais foram menores do que a taxa intrínseca do restaurante (ver figuras 18 e 19). No modelo BCG -1, a taxa final subiu consideravelmente, cerca de 60% (ver figura 20).

Por fim, no último restaurante a evolução foi diferente nos três modelos. No modelo *Random*, a taxa de conversão desceu cerca de 20% (ver figura 21). Já nos modelos CR e BCG as taxas foram ligeiramente inferiores a 14%.

R1 - CR=8%

Random – Prive vs Quality

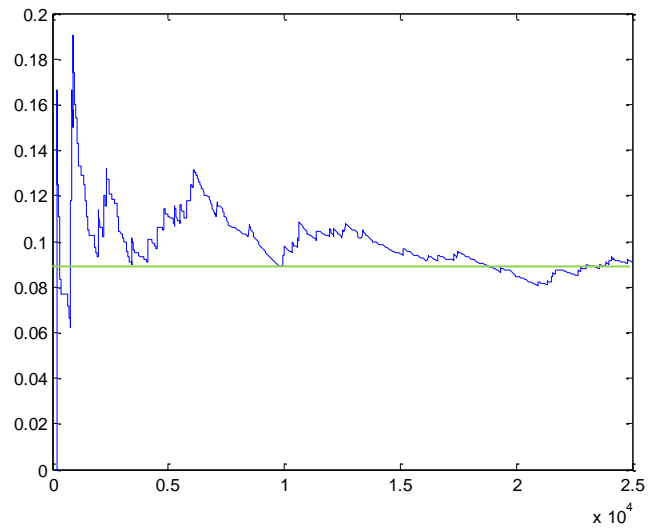


Fig. 16 – Gráfico da taxa de conversão de R1 no modelo *Random*

R1 - CR=8%

CR – Prive vs Quality

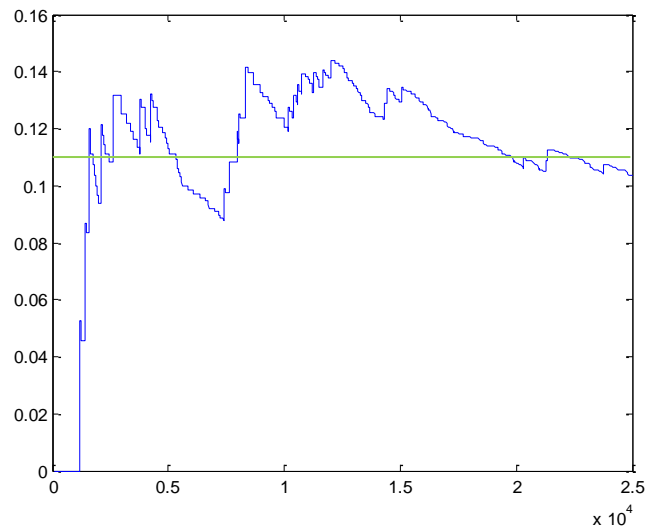


Fig. 17 - Gráfico da taxa de conversão de R1 no modelo *CR*

R1 - CR=8%

BCG 1 – Prive vs Quality

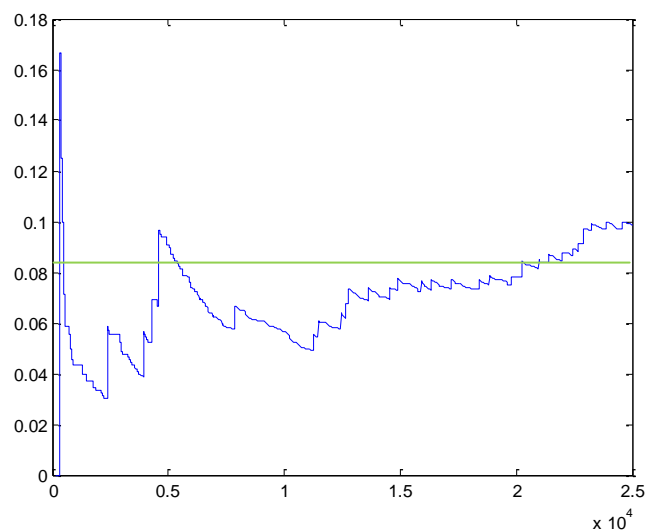


Fig. 18 - Gráfico da taxa de conversão de R1 no modelo *BCG 1*

R2 - CR=3%

Random – Prive vs Quality

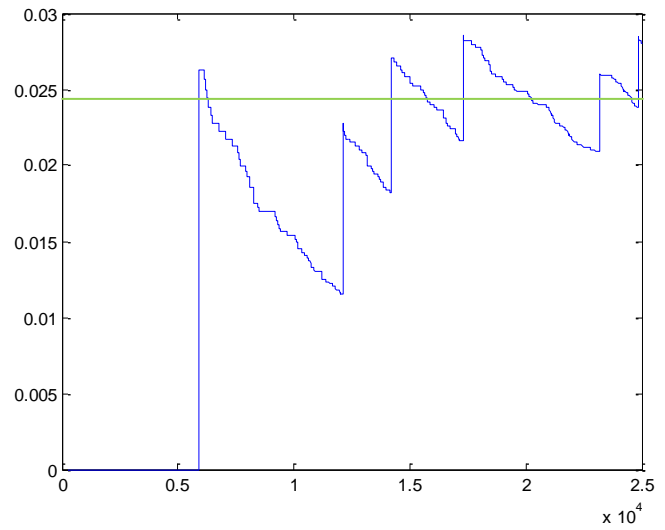


Fig. 19 - Gráfico da taxa de conversão de R2 no modelo *Random*

R2 - CR=3%

CR – Prive vs Quality

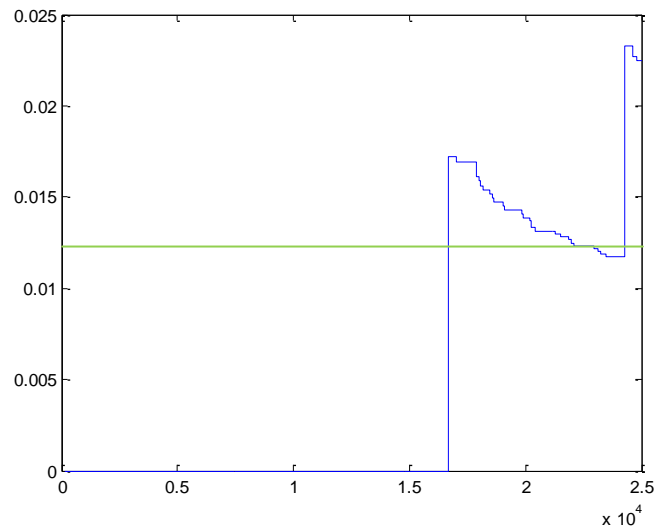


Fig. 20 - Gráfico da taxa de conversão de R2 no modelo *CR*

R2 - CR=3%

BCG 1 – Prive vs Quality

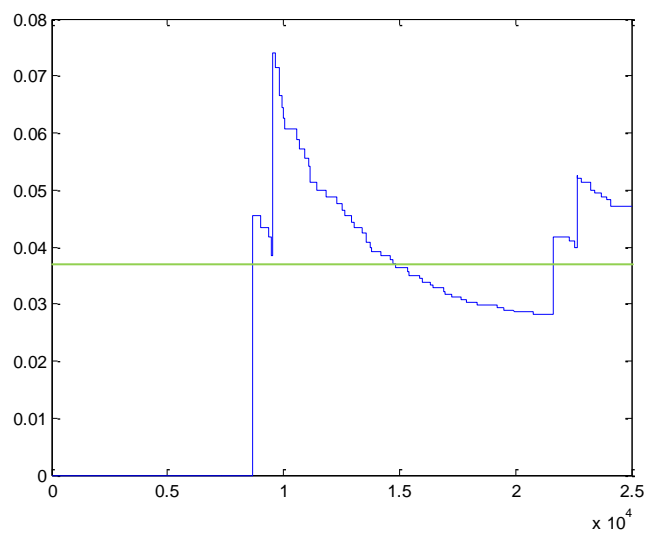


Fig. 21 - Gráfico da taxa de conversão de R2 no modelo *BCG 1*

R3 - CR=14%
Random - Prive vs Quality

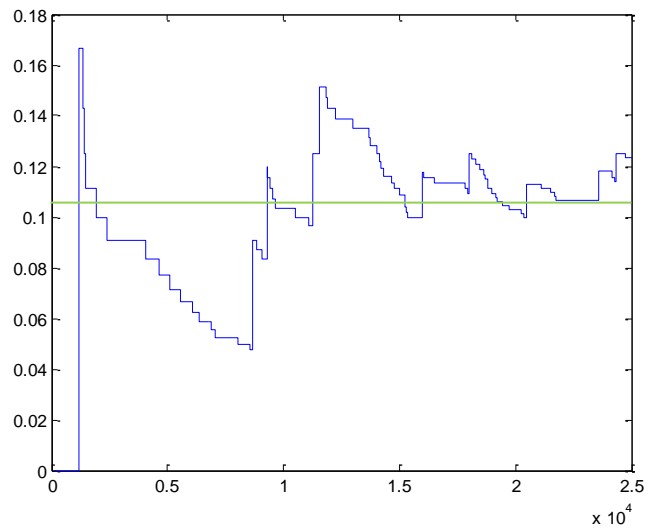


Fig. 22 - Gráfico da taxa de conversão de R3 no modelo *Random*

R3 - CR=14%
CR - Prive vs Quality

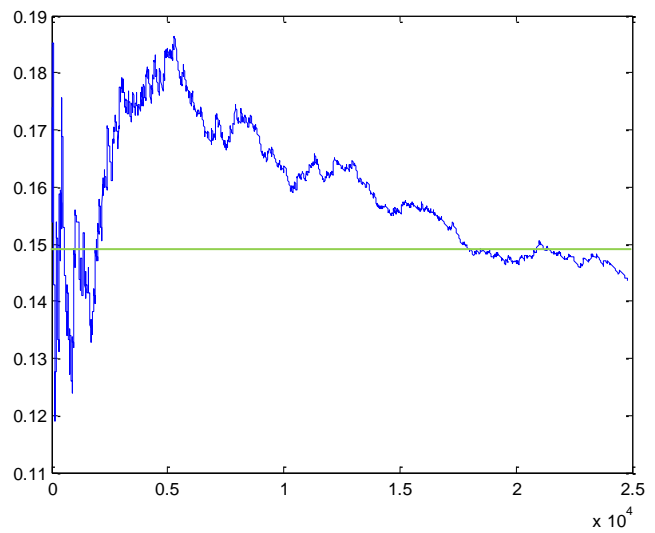


Fig. 23 - Gráfico da taxa de conversão de R3 no modelo *CR*

R3 - CR=14%
BCG 1 - Prive vs Quality

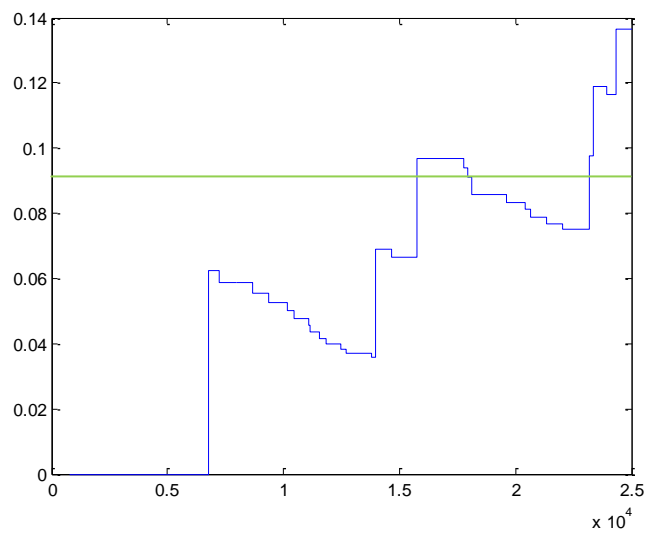


Fig. 24 - Gráfico da taxa de conversão de R3 no modelo *BCG 1*

Capítulo 7

Conclusão e Trabalho Futuro

7.1 Conclusão

Nesta dissertação construímos uma *Framework*, como base de estudo para os modelos de *scoring*, de forma a melhorar a gestão e encaminhamento de leads. Esta plataforma de simulação é aplicável a diversas áreas, apesar deste estudo apenas se centrar na análise de restaurantes e dos seus clientes numa área que representa uma cidade média como Coimbra.

O processo de desenvolvimento da plataforma de simulação passou pela construção de um cenário específico, a partir de valores que se reconheceram como sendo os mais razoáveis e prováveis.

Pela análise dos resultados dos testes do Tipo 1, podemos verificar que os diferentes atributos têm uma importância diferente na selecção dos restaurantes. Qualquer modelo que representa um atributo tem taxas de conversão superiores ao modelo *Random*, onde os atributos não têm qualquer papel na selecção, ou seja, por menor que seja o peso. Contudo, um bom valor num determinado atributo tem uma contribuição positiva na hora da selecção dos restaurantes.

Podemos ainda concluir que os restaurantes cujos atributos se encontram em extremos, nomeadamente o Preço (*Price*) e a Distância (*Distance*) no caso de ser elevada, são muito selectivos, uma vez que a maioria das pessoas procura os valores médios.

Pelos resultados ao nível dos valores do *Review Score*, podemos afirmar que neste estudo estes são independentes do número de conversão. Desta forma, um restaurante que tenha taxas de conversão insatisfatórias, pode apostar na melhoria da qualidade do serviço ou ter preços competitivos, de forma a poder aumentar o seu *Review Score* e futuramente aumentar a sua taxa de conversão ou pelo menos o número de selecções.

Com os resultados dos testes do Tipo 2, podemos concluir que os novos restaurantes, uma vez que, inicialmente têm como valores nos seus atributos as médias intrínsecas dos restaurantes existentes, têm boas probabilidades de serem seleccionados. Porque face aos restaurantes menos

qualificados que já existem, partem de uma posição superior. Este aspecto é positivo, uma vez que, desta forma os novos restaurantes são objecto de seriação ao longo do tempo e quando se obtém a média real após 20 conversões, esta reflecte o real valor do restaurante.

A seriação pode ser vista pelas diferentes taxas reais de conversão que cada um já tem, ou pela tendência que levam ao longo da simulação. De notar que para a maioria dos restaurantes a simulação efectuada de 25 mil clientes, não foi suficiente para as taxas reais poderem ser calculadas.

De facto, as taxas de conversão nos modelos CR e BCG foram inferiores às taxas obtidas pelos mesmos modelos para os 500 restaurantes (Tipo 1). No primeiro modelo a explicação pode ser dada pelo facto de todos os restaurantes inicialmente terem uma taxa de conversão igual a 8% e não existirem restaurantes com médias superiores, levando a taxa de conversão a subir. Já o modelo BCG falha, por não contemplar uma excepção para calcular o score dos novos restaurantes, uma vez que uma das duas principais variáveis de análise deste modelo é o número de conversões.

Finalmente, pelos resultados dos testes do Tipo 3, podemos realçar que as taxas de conversão não são uniformes ao longo da simulação, mesmo perante modelos iguais. Apesar disso, as taxas de conversão tendem a estabilizar a partir do último terço da simulação e aproximarem-se entre si com o final da simulação.

O primeiro fenómeno, a volatilidade da taxa de conversão, ocorre porque os resultados decorrem de processos probabilísticos. No segundo caso, a estabilização deve-se ao facto de ao longo da simulação quer o número de conversões quer o número seleções aumenta e com isso a taxa de conversão vai sendo progressivamente mais estável.

Podemos ainda concluir que a taxa de conversão de um restaurante não pode ser corrigida num curto espaço de tempo e que vai aumentando ou diminuindo lentamente. Constatamos ainda que, restaurantes com uma qualidade mais alta têm tendência a ter taxas de conversão altas e restaurantes de pouca qualidade terem baixas taxas de conversão. Apesar disso, um dos resultados mostra que um dos restaurantes com baixa taxa de conversão, consegue subir substancialmente a sua taxa. Este fenómeno pode ocorrer aquando de uma melhoria significativa da qualidade do restaurante.

As bonificações criadas trouxeram o benefício de simular alguns dos comportamentos humanos que os modelos de *scoring* não poderiam captar. Este aspecto é muito relevante uma vez que neste universo existem sempre aspectos subjectivos que não estão directamente ligados à

qualidade dos prestadores de serviço e muitas vezes os negócios não funcionam de uma forma “justa”.

Em suma, podemos concluir que os modelos de scoring produzem um aumento do número de conversões, sendo este o principal objectivo. Este aumento do número de conversões traduz a viabilidade deste processo, uma vez que, existem benefícios quer para os restaurantes quer para o gestor da plataforma.

O procedimento que envolve o aumento do número de conversões não tem um fim, não é mais do que um processo cíclico na tentativa de aumentar constantemente a percentagem o número de clientes. Um aumento, por marginal que seja, pode significar um aumento considerável na margem de lucro.

7.2 Trabalho Futuro

A construção de um cenário, a partir de fontes reais, traria certamente mais veracidade e fidelidade aos resultados obtidos. No mesmo sentido, seria benéfico fazer levantamento no terreno a partir de respostas dadas pelos clientes e restaurantes, para descortinar o que os clientes procuram num restaurante e o que os restaurantes estão dispostos a fazer para atrair mais clientes.

A plataforma de simulação desenvolvida nesta dissertação tem uma grande margem de melhoramento, uma vez que inicialmente foi pensada como sendo um ponto de partida para a criação de uma plataforma numa linguagem compilada, estando esta preparada para suportar um maior número de clientes e prestadores de serviço e ainda assim, com um melhor comportamento computacional.

Outro passo relevante que pode ser objecto de melhoria é a inclusão das funções que permitem criar o cenário e as que possibilitam estudar os resultados das simulações, na plataforma de simulação. Desta forma o cenário passaria a ser criado e estudado numa única ferramenta, sendo mesmo possível visualizar os resultados em tempo real.

Outras melhorias que tornariam a plataforma de simulação mais completa seriam:

- Tornar a plataforma mais dinâmica para a configuração inicial do cenário;
- Possibilitar que sejam usados dados reais de uma forma mais simples, sem que isso necessite de uma alteração na estrutura da simulação;
- Criar ficheiros de output mais ricos em termos de informação no final de cada simulação.

Bibliografia

- Adriaans, P. and Zantinge, D. (1996), *Data mining*, New York, NY: Addison-Wesley.
- Berry, M. J. A., and Linoff, G. (1997), *Data Mining techniques : For marketing, sales, and customer support*, New York, NY: John Wiley & Sons, Inc.
- Berson, A., Smith, S., and Thearling, K. (1999), *Building Data Mining Applications for CRM*, New York, NY: McGraw-Hill
- Burkey, E., Breakfield, V. (2001), *Designing a Total Data Solution, Technology, Implementation, and Deployment*, Auerbarch
- Capon, N, *Credit Scoring Systems: A Critical Analysis*, *Journal of Marketing*, 46:2 (1982:Spring)
- Carneiro, Ana (2008), *Using Web Data for Measuring the Effectiveness of an E-Commerce Site*
- Dai H, Nie Z, Lee W, Lingzhi Z, Wen J.Rong, Li, Y, (2006), *Detecting Online Commercial Intention (OCI)*
- Frank, I, (2006), *Calculating Geographic Distance: Concepts and Methods*, Canadian Institute for Health Information, Toronto, Ontario, Canada
- Li, K., Efosa, C. Idemudia, Lin, Z, Yu, Y, *A framework for intermediated online targeted advertising with banner ranking mechanism*
- Kotler, P., Bowen, J., and Makens, J. (2006), *Marketing for Hospitality and Tourism (4th Ed)*, Upper Saddle River, NJ: Prentice-Hall, Inc.
- Peppers, D., Rogers, M., and Dorf, B. (1999), *Is your company ready for oneto-one marketing*, *Harvard Business Review*, Jan-Feb, 151-160.
- Silva, Filipe (2009), *Sistema de Medida de Confiança para Selecção de Empresas em Sistemas B2B*.
- Silva, Luís (2009), *API de Serviços Web e Extensão para Software de Webanalytics/Adservering*
- Steven E. Phelan, Markus V. (2003), *The BCG Matrix Revisited: A Computational Approach*, University of Nevada Las Vegas, Università Bocconi

ANEXOS

Código referente à função de cálculo do *Review Score* referida no documento:

```
% Feedback
function [Score] = Feedback(Mean, Profile, RSi)

hasreviewed=rand(1);

if hasreviewed <= 0.15

aux=randi(6);

Random_factor=aux/20;

if Profile==-1
    Score=Mean*(1-Random_factor);
elseif Profile==0
    Score=Mean;
else
    Score=Mean*(1+Random_factor);
end

aux_RSi_index=randi(9);
aux_RSi_index2=randi(29);

Random_factor2=aux_RSi_index/20;
Random_factor3=aux_RSi_index2/20;

if RSi < Mean
Score=Score-(Random_factor3);
else
Score=Score+(Random_factor2);
end

else

    Score=0;

end
```

Modelos de Scoring - Bonificação	NC	NS	NRS	Média CR	Média ICR	CR - ICR	nº de vezes: CR > ICR	Médis RS	Média IRS	RS-IRS	nº de vezes: RS > IRS	CR	Ganho em relação há ICR
Random - Low Cost	10043	121530	1595	8,04	7,99	0,06	225	3,48	3,45	0,02	125	8,26%	3,43%
Random - P vs Q	9798	121687	1529	7,78	7,99	-0,21	209	3,50	3,45	0,04	133	8,05%	0,77%
Random - HQ	9827	121628	1433	8,00	7,99	0,01	204	3,51	3,45	0,05	136	8,08%	1,12%
RS - Low Cost	10369	122384	1524	7,98	7,99	-0,01	239	3,53	3,45	0,08	103	8,47%	6,04%
RS - P vs Q	10229	122216	1558	8,01	7,99	0,02	259	3,56	3,45	0,11	108	8,37%	4,75%
RS - HQ	10180	122284	1515	7,81	7,99	-0,18	243	3,55	3,45	0,10	103	8,32%	4,19%
CR - Low Cost	12836	122311	1949	8,17	7,99	0,18	246	3,53	3,45	0,08	114	10,49%	31,35%
CR - P vs Q	12923	122320	1902	8,09	7,99	0,10	245	3,48	3,45	0,02	99	10,56%	32,23%
CR - HQ	12592	122337	1922	8,00	7,99	0,01	234	3,57	3,45	0,12	115	10,29%	28,82%
Distance - Low Cost	10237	122348	1510	8,06	7,99	0,07	259	3,44	3,45	-0,01	136	8,37%	4,72%
Distance - P vs Q	10630	122275	1595	8,56	7,99	0,57	286	3,46	3,45	0,01	133	8,69%	8,80%
Distance - HQ	10271	122282	1556	8,23	7,99	0,24	264	3,52	3,45	0,06	128	8,40%	5,12%
Price - Low Cost	10204	122315	1520	8,05	7,99	0,06	248	3,49	3,45	0,04	141	8,34%	4,41%
Price - P v Q	10826	122190	1607	8,66	7,99	0,67	303	3,49	3,45	0,04	131	8,86%	10,89%
Price - HQ	10472	122330	1527	8,19	7,99	0,20	277	3,48	3,45	0,03	137	8,56%	7,14%
Mean - Low Cost	11253	122270	1762	8,01	7,99	0,02	230	3,56	3,45	0,10	90	9,20%	15,19%
Mean - P vs Q	12028	122268	1790	8,52	7,99	0,53	280	3,48	3,45	0,03	105	9,84%	23,12%
Mean - HQ	11539	122310	1688	8,24	7,99	0,25	259	3,59	3,45	0,14	94	9,43%	18,08%
Top Average - Low Cost	10505	122247	1578	8,22	7,99	0,23	264	3,59	3,45	0,14	98	8,59%	7,55%
Top Average - P v Q	10687	122275	1589	8,29	7,99	0,30	266	3,60	3,45	0,15	108	8,74%	9,39%
Top Average - HQ	10523	122261	1611	8,11	7,99	0,12	262	3,61	3,45	0,15	104	8,61%	7,72%
BCG - 1 - Low Cost	11923	122352	1837	8,27	7,99	0,28	264	3,53	3,45	0,07	112	9,74%	21,96%
BCG - 1 - P vQ	12486	122264	1879	8,62	7,99	0,63	274	3,51	3,45	0,06	112	10,21%	27,81%
BCG - 1 - HQ	12131	122266	1821	8,12	7,99	0,13	256	3,53	3,45	0,08	126	9,92%	24,18%

Tabela 14 – Tabela completa dos resultados referentes aos testes do Tipo 1