

Using Web Data for Measuring the Effectiveness of an E-Commerce Site

Ana Carneiro

Dissertação para a obtenção do grau de
Mestre em Análise de Dados e Sistemas de Apoio à Decisão

Orientadores:

Prof. Doutor Alípio Jorge

Prof. Doutor Pedro Quelhas Brito

Outubro de 2008

Biographic Note

Ana Isabel Ribeiro Carneiro was born in Marco de Canaveses, Portugal. She studied and lived in her hometown until 1999, when she joined the University of Porto, moving to the same city. In 2004 completed a degree in Economics at the Faculty of Economics of the University of Porto. In the same year began her career in the Grupo MCoutinho. In 2005 wins the contest “Youth Values of Economy” of the Order of Economists, in the specialty of Political Economy, being co-author of the work "The behaviour of economic agents in a context of non-observed economy". In 2006 enters the Master in Data Analysis and Decision Support Systems. She moves to Lisbon, in the year of 2008, to work in the Office for Planning, Strategy, Appraisal and International Relations of the Ministry of Finance and Public Administration. In the same year is invited to rejoin the Grupo MCoutinho to start working as a financial controller.

Acknowledgements

It is the time to thank everyone who helped me to accomplish this step of my life.

First I would like to thank my dissertation supervisors: Doctor Alípio Mário Jorge and Doctor Pedro Quelhas Brito. Doctor Alípio Mário Jorge provided me with the necessary understanding of my forwards and backs. Doctor Pedro Quelhas Brito provided me the necessary enthusiasm to do better.

I would like to thank Introduxi that gave the data. A special thanks to Timóteo Moreira.

I would like to thank Marcos Domingues for housing Introduxi data in his data warehouse. I would like to thank LIADD for pre-processing data.

I would like to thank all my tutors in my different jobs. They provided me with a practical experience that was of great importance for this dissertation.

I would like to thank my husband, parents and friends. They have been always supportive and loving in the most difficult moments.

Finally I would like to say that the development of this dissertation was supported by the POSC/EIA/58367/2004/Site-o-Matic Project (Fundação Ciência e Tecnologia), FEDER e Programa de Financiamento Plurianual de Unidades de I & D.

Abstract

An electronic commerce web site is successful if it achieves the purpose why it was created. Being successful is a necessary condition for the continuity of the business. For a company owning an e-commerce site it is not enough to know that they are selling and making a profit out of their activity; it is important to understand how they are doing it and how they can improve in order to do better and fight the competition. In this dissertation we apply the measure of the success of a web site to a real case, determining the pages which are not so successful and need a little more attention. We then determine the success measures within the customer sessions, understanding what are the more problematic pages among the buying customers. We segment the customers according to sales and navigation behaviour; this allows us to define customized marketing approaches. We make a comparison between the sales data and the success data; subsequently we determine outliers.

Um site de comércio electrónico tem sucesso se atinge o objectivo para o qual foi criado. Ter sucesso é uma condição necessária para a continuidade do negócio. Para uma empresa que possui um site de comércio electrónico não é suficiente saber que estão a vender e a lucrar com a sua actividade, é também importante compreender como é que o estão a fazer e como é que se podem modernizar no sentido de fazerem face à concorrência. Nesta dissertação aplicamos a medição do sucesso de um site a um caso real, determinando as páginas que não têm tanto sucesso e que precisam de um pouco de mais atenção. Determinamos também as medidas de sucesso entre as sessões dos clientes, tentando compreender quais são as páginas mais problemáticas entre os clientes que compram produtos no site. Segmentamos ainda os clientes de acordo com vendas e comportamento de navegação; isto permite-nos definir abordagens de marketing personalizadas. Fazemos ainda uma comparação entre dados de vendas e medidas de sucesso, determinando de seguida os outliers.

Index

Biographic Note	i
Acknowledgements.....	ii
Abstract	iii
Index	iv
1 Introduction	1
2 Data the Web and E-Commerce	4
2.1 The role of the Electronic Commerce in the Web	4
2.2 Electronic commerce in a marketing perspective	4
2.2.1 Categories of e-commerce	5
2.2.2 E-marketing planning	7
2.2.3 Marketing Mix and the World Wide Web	8
2.2.4 CRM (Customer relationship management) and PRM (Partner Relationship Management)	10
2.2.5 Mass Marketing, Market Segmentation and One-to-One Marketing	11
2.2.6 The Costumer and E-Loyalty.....	12
2.2.7 Database Marketing and Business Intelligence	14
2.2.8 Online marketing communications	16
2.3 Data mining.....	18
2.3.1 Web Mining	21
2.3.2 Web mining tools.....	22
2.4 Data.....	23
2.4.1 How to collect data?	23
2.4.2 What sources should we use to collect data?	24
2.4.3 How to generate knowledge from data?	27
Preprocessing	27
2.5 How to measure the success of a site?	33
3 The Problem.....	35
3.1 CRISP	35
3.2 Business Understanding.....	36
3.3 Data understanding	37
3.4 Data preparation.....	38
4 Working on the Problem	39
4.1 Modelling.....	39
4.2 Exploiting Introduxi's data	40
4.2.1 The buying process	43
4.2.2 Analysis of selling data.....	45
Analysing turnover	46
Analysing quantities	48
Turnover versus Quantities	50

4.2.3	Log analysis	51
4.2.4	Log analysis - Customer Session	59
5	Evaluation and Deployment Plan	64
5.1	Business Intelligence Analysis	64
5.2	Market Segmentation	70
5.3	Combining sales and success metrics	80
5.4	Deployment Plan.....	81
6	Conclusions	83
6.1	Preliminary Analysis.....	83
6.2	Contact Efficiency	84
6.3	Conversion Efficiency	84
6.4	Focusing on top costumers	85
6.5	Market Segmentation	86
6.6	Site's success per segment.....	87
6.7	Sales versus Efficiency	88
6.8	Main contributions	88
6.9	Future work.....	89
	Bibliography	90
	Appendix 1	93
	Appendix 2	94
	Appendix 3	97
	Appendix 4	107

Table Index

Table 1: List of products by type.	37
Table 2: List of products by family.	38
Table 3: The two different ways of completing a buying process in the Introduxi's website	44
Table 4: Efficiency results per product type considering the entire log	59
Table 5: Efficiency results per product type considering customer session	63
Table 6: Ten customers that bought more quantities of products regardless of the product.	64
Table 7: Ten customers with the highest share of customer.	65
Table 8: Confronting performance navigation results between a top share customer and a top buyer customer	67
Table 9: Average results per cluster considering the variables chosen for market segmentation	73
Table 10: Relative Contact Efficiency results per cluster and by type of product	76
Table 11: Conversion Efficiency results per cluster and by type of product.....	78

Chart Index

Chart 1: How the turnover from the Net is split into types of products.	46
Chart 2: How the turnover from the “classic channel” is split into types of products....	47
Chart 3: How the turnover from the Net is split into families of products.	47
Chart 4: How the turnover from the “classic channel” is split into families of products.	48
Chart 5: How the quantities are split into types of products.....	49
Chart 6: How the quantities are split into families of products.	49
Chart 7: Comparison between the number of quantities sold and the average price by family of product.	50
Chart 8: Confrontation between the weight of quantities (%) and the weight of turnover (%) for families of products.....	50
Chart 9: Page request distribution over the observation period.....	51
Chart 10: Number of sessions over the observation period.....	52
Chart 11: Page request distribution over the week for the observation period.....	52
Chart 12: Orders distribution over the week for the observation period.	53
Chart 13: Page request distribution over the day for the observation period	53
Chart 14: Orders distribution over the day for the observation period.....	54
Chart 15: Top ten clickers.....	54
Chart 16: Top ten buyers.	55
Chart 17: The relative importance of each product for the costumer VGER in terms of value.....	66
Chart 18: The relative importance of each product for the costumer MBIT in terms of value.....	66
Chart 19: Number of clicks per product for the costumer VGER	68
Chart 20: Number of clicks per product for the costumer MBIT	68
Chart 21: Average results per cluster considering the variables chosen for market segmentation	75
Chart 22: Combining sales and success metrics per family of products	80

Figure Index

Figure 1: Four Categories of E-commerce.	6
Figure 2: SOSTACTM – a generic framework for e-marketing planning.	7
Figure 3: The Traditional Consumer Purchase Decision Cycle.	9
Figure 4: The New Consumer Purchase Decision Cycle.....	9
Figure 5: Three phases of customer relationship management.....	11
Figure 6: Phases of the CRISP-DM Reference Model	20
Figure 7: Tasks and Outputs of CRISP-DM Reference Model	20
Figure 8: Server Log File Formats.....	26
Figure 9: Homepage of Introduxi’s website	41
Figure 10: Step 1 of the buying process.	44
Figure 11: Step 2 of the buying process.	44
Figure 12: Step 3 of the buying process.	45
Figure 13: Step 4 of the buying process.	45
Figure 14: Step 5 of the buying process.	45

1 Introduction

The need to be connected by nets goes way back to the beginning of the civilization. Man has always felt the need to interact with others. Throughout history, several ways of interacting have been created, used, developed, maintained or forgotten. Internet came as a new way of being connected to other human beings, to news and all kinds of information.

Some people say Internet is the best thing ever happened, others that it is the devil in disguise. It is not our purpose to judge. Either way, we might say that Internet came as a huge revolution in the way people interact, as well as in the way information is made available.

Internet has also come as a new channel for business. Buying, selling and exchanging products, services or information is commonly done by the Internet channel. This leads us to another known truth about the human being: business means competition. Electronic commerce needs to be analysed, dissected to understand which websites are more popular and why; which sites are more profitable and how.

The interest in electronic commerce websites started flourishing years ago. We still can find websites in different stages of improvement. There is however a common characteristic to all of them, they wish to fulfil the purposes they were conceived for. How can we measure this? How to determine if your e-commerce site is effectively accomplishing the reasons behind its creation? Data mining allows measuring and qualitatively determining answers to these questions. Data mining methods combined with a marketing perspective seems to be a winning team when it comes to measuring the effectiveness of an e-commerce site.

Problem under study

Currently there are several ways for measuring the success of e-commerce sites. There are many companies specialized in e-commerce analysis, each one of those use different

methods and techniques to solve the same problem: Is this website successful? Why or why not? How do you make it successful?

To start looking for answers we should take the following aspects into consideration: which data are we going to use; how we are going to clean the data; where are we going to store the data; what method will be used; which techniques will be applied; what perspective will be used in the interpretation of the results; how might this results be modelled to improve the company under analysis.

We should also not to forget when analysing the results if there are pages that need more attention than others, that need a little more help to be successful, if all the customers behave alike and if the success results are the same for all, if the buying preferences reflect the navigation performance behaviour or understanding to what extent the sales data are in line with the success measures of the web site

Goals

In this dissertation we want to study the effectiveness of an e-commerce site, using web server log files data and commercial data. The idea is to combine web data and selling data to determine the ability of the web site in achieving the goal for which it was created: selling. For this it will be used some Data Mining techniques such as clustering. We will also obtain values for site success metrics following the work developed by Spiliopoulou, on this subject.

The final goal is to provide the company under analysis with some information of value not only concerning the interpretation of the results but also delivering some advice. These advices are supposed to combine technical web data analysis and marketing analysis.

Structure of the dissertation

This dissertation is divided in three parts. In the first part we explain the theoretic concepts underlying this subject. In the second part we analyse, prepare, decide which model we will apply to the data and complete the application. In the third part we evaluate and interpret the discovered results and finally discuss the deployment plan.

In the first part we explain the theoretic concepts underlying this subject. This chapter addresses the marketing concepts necessary to enhance the profitability view of the analysis; details on the concept of data mining; and finally is made a detailed approach on data and how it can be used to measure the success of a site.

In the second part we present the problem and the company behind this case-study. We define the methodology used for conducting the data mining process. We model and evaluate the company's data also some actions are proposed in order to overcome bottlenecks of the process. We analyse how the company's website buying process is organized; evaluate selling results; measure web server log results; propose a business intelligence approach to the website and perform a clustering analysis to the data, creating a new database from selling data and log data. The website success is then measured by each cluster previously determined.

In the third part we evaluate and interpret the discovered results and finally discuss the deployment plan.

2 Data the Web and E-Commerce

The ease and speed with which information exchange and business transactions can be carried out over the Web has been a key driving force in the rapid growth of the Web and electronic commerce (Norguet et al., 2006).

The beginning of the Internet use for commercial purposes can be set in 1993, with the development of the World Wide Web, which is based on the HTML language (HyperText Mark-up Language). The HTML language changed the Internet into an easy tool for researching information. Internet became a global network of interconnected nets (Hortinha, 2001, pp. 39-40).

2.1 The role of the Electronic Commerce in the Web

Electronic commerce may be defined as the process of buying, selling, or exchanging products, services, or information via computer networks (Turban et al., 2008, pp. 4). E-commerce has definitely been changing the world in the way people interact, schedule their days, in the way companies reorganize their selling processes and human resources, in the way government relates to the people, to companies and to other countries. Economy, markets, society, labour market, industry have all been and still are being shaken by e-commerce.

There is a difference in the concepts of e-commerce and e-business that is better to clarify. E-business is a broader concept than EC, not just the buying and selling of goods and services but also servicing costumers, collaborating with business partners, conducting e-learning, and conducting electronic transactions within an organization (Turban et al., 2008, pp. 4).

2.2 Electronic commerce in a marketing perspective

“The ability to track user’s browsing behaviour down to individual mouse clicks has brought the vendor and end costumer closer than ever before. It is now possible for a

vendor to personalize his product message for individual customers at a massive scale, a phenomenon that is being referred to as mass customization.” (Srivastava et al., 2000)

Why is electronic commerce such a mediated subject? Simply because it can be an huge source of income. Scientists like Spiliopoulou (Spiliopoulou et al., 2001), Srivastava and Cooley (Srivastava et al., 2000), Fayyad (Fayyad et al., 1996a), (...) write about the importance of web usage mining and how it can contribute to optimize the companies’ web sites. They recognize that the reason behind that importance is the money it can generate. For marketers it is clear that it is all about the return on investment. For Travis (2003, pp. 5) approaching the World Wide Web in the right way brings considerable advantages; “There are four key benefits from a customer-centred approach: higher revenues, loyal customers, improved brand volume and process improvement.”

The electronic commerce is, nowadays, a major distribution channel for goods and services. E-commerce has changed the way economic agents and markets interact, it has changed the flow of products and services, it has changed the way labour market performs, it has changed consumers’ behaviour and values. Accordingly to Turban (2008, pp. 4) electronic commerce (EC) is the process of buying, selling, or exchanging products, services, or information via computer networks.

Another important concept to retain is the e-marketing concept, which implies achieving marketing objectives through use of electronic communications technology. Electronic marketing is a subset of electronic business (Chaffey, 2004, pp. 318).

2.2.1 Categories of e-commerce

Accordingly to Rayport et al. (2003, pp. 4) there are four categories of e-commerce: business-to-business, business-to-consumer, peer-to-peer, and consumer-to-business. Rayport et al. (2003, pp. 4, 5) defines each category as follows:

		Business originating from...	
		<i>Business</i>	<i>Consumers</i>
And selling to...	<i>Business</i>	B2B	C2B
	<i>Consumers</i>	B2C	P2P

Source: (Rayport et al., 2003).

Figure 1: Four Categories of E-commerce.

- Business-to-Business (B2B) activity refers to the full spectrum of e-commerce that can occur between two organizations. Among other activities, this includes purchasing and procurement, supplier management, inventory management, channel management, sales activities, payment management, and service and support.
- Business-to-Consumer (B2C) e-commerce refers to exchanges between businesses and consumers.
- Peer-to-Peer (P2P) exchanges involve transactions between and among consumers. These exchanges can include third-party involvement.
- Consumer-to-Business (C2B), consumers can band together to present themselves as a buyer group in a C2B relationship. These groups may be economically motivated or socially oriented.

2.2.2 E-marketing planning

Traditional marketing implies the definition of a strategy to be implemented in the company (Kotler, 1997). Accordingly to Chaffey (2004, pp. 320) an e-marketing plan is needed in addition to a broader e-business strategy to detail how the objectives of the e-business strategy will be achieved through marketing activities such as marketing research and marketing communications. The stages involved are presented in figure 2 and can be summarized as:

- Situation – where we are now?
- Objectives – where do we want to be?
- Strategy – how do we get there?
- Tactics – how exactly do we get there?
- Action – what is our plan?
- Control – did we get there?



Source: (Chaffey, 2004)

Figure 2: SOSTACTM – a generic framework for e-marketing planning.

The loop is closed because by using the analysis of metrics collected as part of the control stage it is possible to continuously improve e-marketing through making enhancements to the web site and associated communications.

2.2.3 Marketing Mix and the World Wide Web

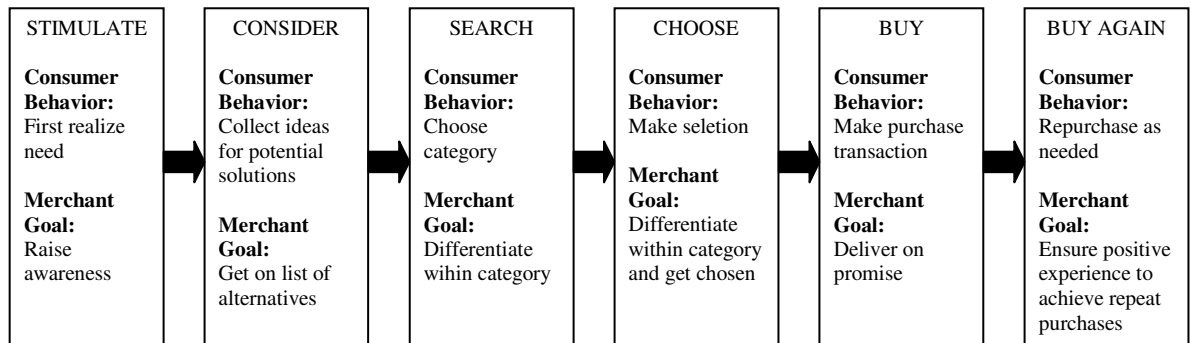
Marketing tactics to implement strategies and objectives are traditionally based around the elements of the marketing mix. The marketing mix - the 4Ps of Product, Price, Place and Promotion – is used as an essential part of implementing marketing strategy by many practitioners. (Chaffey, 2004, pp. 345)

Accordingly to Reis (2000, pp. 84) with the changes caused by the technological revolution it was necessary to consider, to develop and extend this concept of marketing mix. This concept of marketing mix tends to lead to a product rather than a customer orientation – a key concept of market orientation and indeed a key internet marketing concept. (Chaffey, 2004, pp. 346)

The online buying process

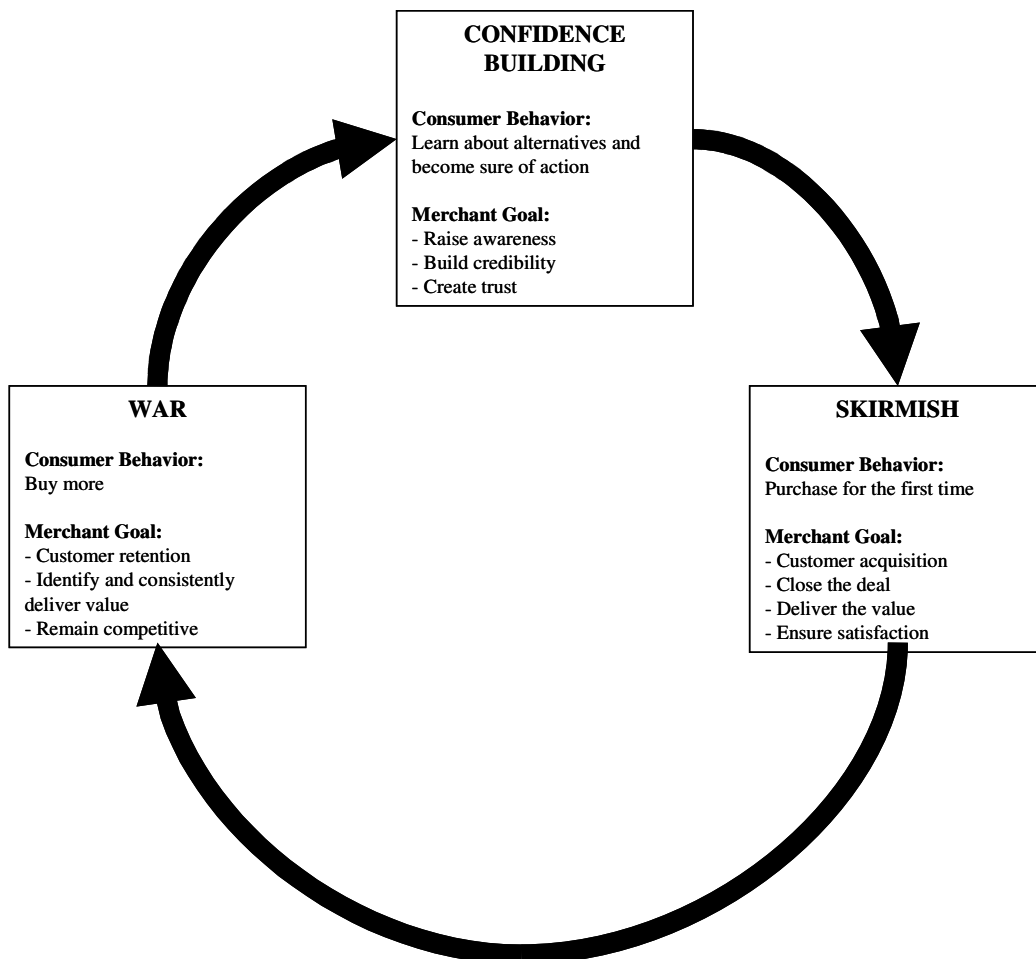
The traditional marketing model has been helpful when developing strategies and programs because it clarifies the consumer actions to motivate throughout the decision-making process. (Windham et al., 2000, pp. 80) There are basically six stages in an individual's purchase decision cycle as can be seen in figure 3.

The web context has a different Consumer Purchase Decision Cycle as can be seen in figure 4. However there are similarities between these two models. "The consumers are generally the same; they have the same needs with the same amount of money to spend. But in the Web World, the decision cycle becomes more compressed because information is so readily available. The consumer is empowered and is often reached via new marketing vehicles. The steps blend together into fewer discrete phases." Windham et al. (2000, pp. 82)



Source: (Windham et al., 2000)

Figure 3: The Traditional Consumer Purchase Decision Cycle.



Source: (Windham et al., 2000)

Figure 4: The New Consumer Purchase Decision Cycle.

2.2.4 CRM (Customer relationship management) and PRM (Partner Relationship Management)

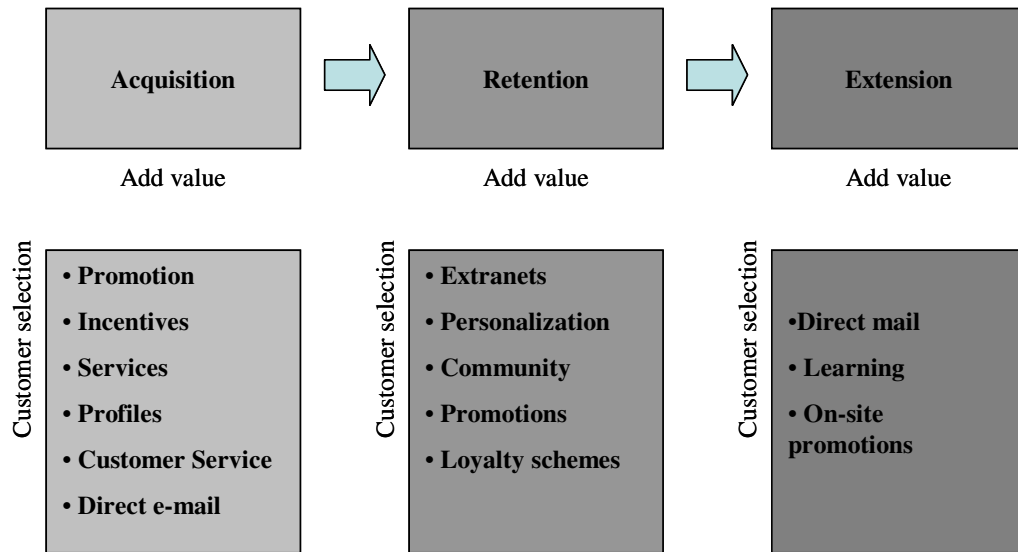
Customer relationship management (CRM) focuses on building long-term and sustainable customer relationships that add value both for the customer and the selling company (Turban, 2008, p. 606). In figure 5 it can be seen the “Three phases of customer relationship management” (Chaffey, 2004).

Accordingly to Rayport (2003, pp.249-250) CRM “refers to technology systems that help the company serve, satisfy, and retain costumers. CRM systems help companies store data about their customers (including who they are and in what ways they have interacted with the company), and they help employees to better communicate and serve costumers.” However the internet has changed CRM. With internet there are more disposable data, in real time.

In Business-to-Business (B2B) category of electronic commerce, CRM is known as PRM (Partner Relationship Management). The Partner Relationship Management (PRM) is a business strategy that focuses on providing comprehensive quality service to business partners (Turban, 2008). According to Turban (2008, p. 280) many customer service features of B2C are also used in B2B. Nevertheless, corporate costumers may require additional services.

Internet may seem an anonymous space but in fact it is easier to “track customers, their purchase histories, and their preferences online than in traditional business setting.” (Reichfeld, 2000) Every click made by a customer is recorded in the web server of the web site.

There are however some difficulties for implementing CRM systems. CRM systems are usually not set up solely for costumer service representatives. They are integrated throughout the company. CRM programs can positively affect a company’s relationship with its costumers but it is essential that the application is implemented properly for the company to realize the full potential of the system. (Rayport, 2003, pp.250)



Source: (Chaffey, 2004)

Figure 5: Three phases of customer relationship management.

2.2.5 Mass Marketing, Market Segmentation and One-to-One Marketing

Since the beginning of market analysis, numerous theories have been developed about how to group consumers into addressable segments based on common attributes so that strategies and programs could be developed to effectively market to them. (Windham et al., 2000, pp. 58)

Mass Marketing, Market Segmentation and One-to-One Marketing are three different approaches used to reach customers. Turban et al. (2008, pp. 162-163) defines each approach as follows:

Mass Marketing → embraces marketing efforts targeted to everyone (the “masses”)

Market Segmentation → refers to the practice of promoting a product or service to a subset of customers or prospects

One-to-One Marketing → one-to-one means not only communicating with customers as individuals, but possibly developing custom products and tailored messages based on the customer's spoken and unspoken needs

There is one important subject when it comes to detailed customer segmentation: privacy. (Seybold, 2002) On the one hand customers require personalized information, targeted offers and special levels of service. On the other hand customers wish that the company shares its segmentation models and assumptions with them; customers wish to know which category they belong to, what are the benefits of being in that category and what it can be done to change segment. Sharing customers' information with third parties without their explicit consent may be unethical or even illegal.

There are some statistical methods for conducting market segmentation. In chapter 2.2.7 Database Marketing and Business Intelligence we will describe statistical methods for doing market segmentation.

2.2.6 The Customer and E-Loyalty

Understanding the consumer behaviour in the World Wide Web is absolutely essential for the success of the business. These way marketers will be able to respond to their consumers' needs on time. After all, competitor suppliers are just one click away and as Reichheld (2000) states, customers' tolerance for inconsistency and mediocrity is rapidly disappearing.

Turban et al. (2008, pp.167) defines e-loyalty as the customer's loyalty to an entity that sells online, being it apparel, music, books or any kind of service. Customer acquisition and retention is a critical success factor in e-commerce. Liao et al. (2006) advises that the maturity and low cost of technology in this business lowers the entrance threshold of new competitors, moreover the transparency of information makes the business model easily imitable by competitors.

In e-commerce the cost of acquiring a customer is often higher than in traditional retail channel. Reichheld (Reichheld, 2000) analyzed customer life-cycle economics in several e-commerce sectors. He found that the general pattern - early losses, followed by rising profits - is actually exaggerated on the Internet. In apparel e-tailing, for example, new customers cost 20% to 40% more for pure-play Internet companies than for traditional retailers with both physical and on-line stores. But, also according to the same author, there is evidence that when a relationship consolidates the customers tend to focus in one supplier. This seems to happen more in the business-to-business sector.

After acquiring a new customer how to gain his loyalty? "To gain the loyalty of customers, you must first gain their trust." (Reichheld, 2000) Accordingly to Liao et al. (2006) to maintain customers' trust you have to constantly improve the usability of your web site. The web site is the only way the customer has to get to know its supplier. Therefore the sites' usability is of extreme importance when it takes to decide whether to trust or not in the supplier.

The basis for loyalty is not technological; it is based on old-fashioned customer service basics like: quality customer support, on-time delivery, compelling product presentations, convenient and reasonably priced shipping and handling, and clear and trustworthy privacy policies. What is actually changing is the rhythm at which economic are played out, and the need for speed in improving products and services. (Reichheld, 2000)

Companies must constantly deliver to their customers a total customer experience, Seybold (2002, pp. 108) defines "total customer experience" as: "A consistent representation and flawless execution, across distribution channels and interaction touchpoints, of the emotional connection and relationship you want your customers to have with your brand."

When customers identify themselves with a brand it is more likely that they become and remain loyal to it. Every time they come into contact with the brand if they receive positive experiences they are reinforcing that loyalty. (Seybold, 2002, pp. 107)

A company must define what kind of costumers they are willing to attract because the site design strongly defines which type of customers will be attracted by the web site. When defining what kind of costumers to attract and which ones to avoid the company must be aware of the different categories of on-line customers. There are some types of loyalty-oriented customers and some types of customers who flit – like butterflies - from site to site seeking bargains. (Reichheld, 2000)

For companies that use both channels (the traditional and the web) for business it is important to balance both in terms of human resources. A company should not view the web channel as a mere way to reduce costs by bypassing its commissioned sales force. Reichheld (2000) gives the example of a successful company who seamlessly integrates its web channel with its traditional channel. This company pays sales commission independently of the channel which was used to sell, because this way the sales representatives direct customers to the most convenient channel.

Windham et al. (2000, pp. 70) concludes that if web retailers wish to establish web brand loyalty and remain competitive they must provide those components that create a good consumer experience.

2.2.7 Database Marketing and Business Intelligence

When the scale of data manipulation, exploration, and inference grows beyond human capacities, people look to computer technology to automate the bookkeeping. (Fayyad et al., 1996a)

People who devote themselves to online market research find that too much data is available. That might be a problem when it comes to extracting knowledge from such a big amount of data. To use data properly, one needs to organize, edit, condense, and summarize it. This process is extremely important for supporting business decision making. However such a task may be expensive and time consuming. The solution to this problem is to automate the process by using data warehousing and data mining. The

process of extracting useful information from databases in order to support better business decision making is known as business intelligence. (Turban et al., 2008, pp. 178)

Companies which possess thousands of facts about their customers desire to be able to project personalised information and services to customers on the exact time those are required. "Database marketing came about when marketers realized that they could use the power of modern computer technology to begin one-to-one dialogs with thousands of customers, to learn what they want and supply them with individualized products and services that meet their needs. This process has been tremendously advanced by the advent of the Internet." (Hughes, 1994, pp. 27)

Clustering

In order to perform a clustering analysis, it is necessary to dominate certain statistical knowledge. Often, before creating clusters from the disposable data it is necessary to check on the correlation of the variables and reduce the number of variables that will be used for clustering. Next we will briefly explain some of the concepts involved in clustering.

The Principal Components Analysis is a technique of exploratory multivariate analysis that transforms a set of correlated variables in a smaller set of independent variables, linear combinations of the original variables, known as "principal components" (Maroco, 2003, pp. 231).

The Principal Components Analysis is seen as reduction data method. It enables the summarization of information of several correlated variables (and in a certain way redundant) in one or more independent linear combinations which represent the huge majority of the information contained in the original variables. (Reis, 1997)

The Principal Components Analysis gives information about the percentage of the variance explained by each component, being that the first component explains the

maximum of the variance of the data. The last component will be the one with the least contribution for the total variance of the data. (Reis, 1997)

The analysis of groups or "clusters" is a technique of exploratory multivariate analysis that can group individuals or variables in homogeneous or compact groups for one or more common characteristics. Each observation belonging to a particular cluster is similar to all the others belonging to that cluster, and is different from the observations belonging to the other clusters (Maroco, 2003, pp. 295).

There are several methods of classificatory analysis, being the most common the hierarchical methods which construct a hierarchy of partitions. This technique is based, generally, in a matrix of distances (or similarities) calculated between the different variables or cases. (Reis, 1997)

Clustering using “SPAD Data Analysis”

The software that was chosen for doing Clustering Analysis in this dissertation was “SPAD Data Analysis”. SPAD was chosen because it was available at the university.

SPAD is a company which produces software for Data Quality Management, Data Mining and Predictive Data Analysis. SPAD aims to help their customers to better take their decisions. SPAD business intelligence products help companies to turn large amount of data into useful and strategic information for business processes and performance. (www.datasoft.com)

2.2.8 Online marketing communications

The way a company interacts with its costumers is of extreme importance both for being noticed but also for the customers to feel that they are important to the company. Chaffey (2004, pp. 380-385) in his book: E-Business and E-Commerce Management, outlines five of the main methods of online marketing communications.

- ❖ Search engine registration: Search engines are used to look for information whether it is about a company or its products. If an organization is not registered in with a search engine, then the traffic volume will not be the desirable.
- ❖ Link building: When there is a great number of links to a web site it is more likely that this web site will receive a greater volume of traffic. Efforts to increase the number of links are sometimes referred to as 'link building campaigns'.
- ❖ Viral marketing: This method harnesses the network effect of the Internet and can be effective in reaching a large number of people rapidly in the same way as a computer virus can affect many machines around the world. The speed of transmission and impact of the message must be balanced by the naturally negative perceptions of viruses.
- ❖ Online advertising: A visitor who clicks on banner advertising on an advertising site is then transported through to the destination site of the company who paid for the banner advertising. Banner advertising is purchased for a specific period. Not all companies are interested in this kind of online mean of communication.
- ❖ E-mail marketing: E-mail is a significant communications medium since it is widely used. Some companies receive a considerable revenue from their bet in e-mail marketing campaigns.

It is important to notice that the effectiveness and efficiency of each kind of these five online marketing communications must be carefully followed. The investment in the new technologies does not mean for itself interesting revenues.

“If a Web company is spending most of its marketing dollars on indiscriminate banner ads and on-line coupons, with little investment in building communities and promoting referrals, it is probably building long-term losses into its customer base.” (Reichfeld, 2000)

Sterne (2002, pp.3) warns that “the dot-com bubble has burst.” According to him the web has got value, “but where exactly do we find it? What do we measure?”.

2.3 Data mining

Data mining might be shortly defined as the process of extracting knowledge/ relevant information from large amounts of data. Data mining is applied in many different areas. Secret services use Data Mining to predict terrorist attacks, companies use it in their huge databases for retrieving useful knowledge about their clients, financial analysts use it trying to find patterns in exchange, or the interest rate, and science uses it for improving diagnostics (Fayyad et al., 1996b).

According to Fayyad et al (Fayyad et al., 1996a) “...there is no universally best data mining method; choosing a particular algorithm for a particular application is something of an art. In practice, a large portion of the applications effort can go into properly formulating the problem (asking the right question) rather than into optimizing the algorithmic details of a particular data mining method.”

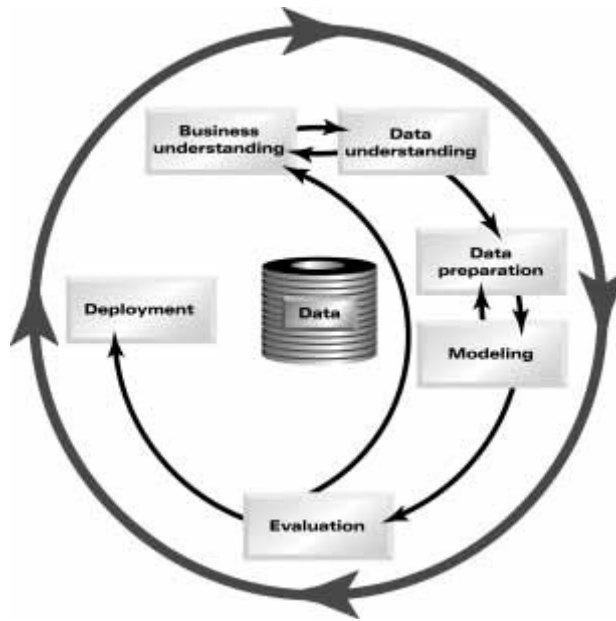
The job of transforming data into information (useful knowledge) takes the use of data mining techniques like for tasks like: association rules discovery, sequential pattern discovery, clustering and others. The knowledge discovery techniques will be explained in chapter 2.4.3 How to generate knowledge from data?, subtitle Knowledge discovery techniques/Pattern discovery.

The success of a data mining project depends on the development and application of a standard process. CRISP-DM provides a generic basis for such a process that has been tried and tested in industry. (Anand et al., 2004)

According to Shearer (Shearer, 2000) CRISP-DM (Cross Industry Standard Process for Data Mining) methodology, as can be seen in figure 6, organizes the process of data mining in six big phases:

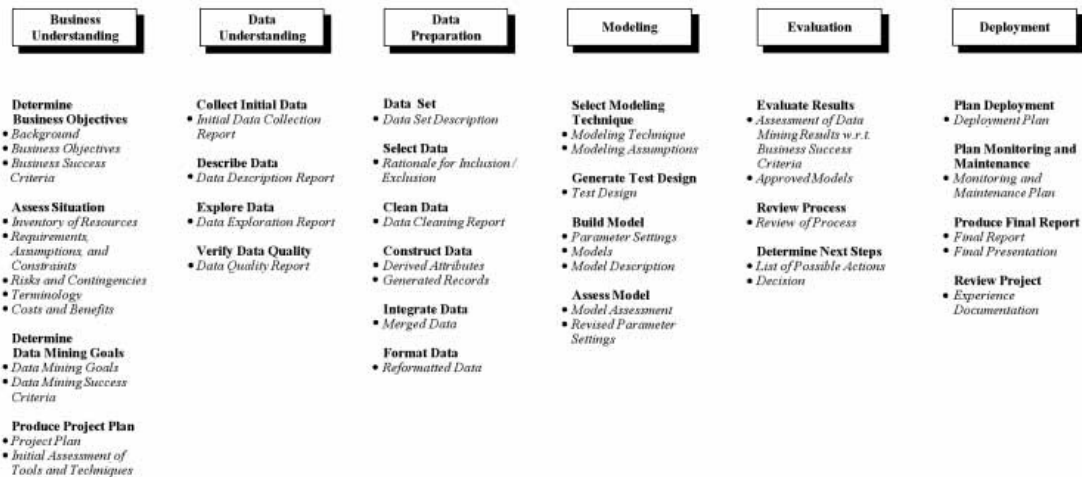
- ❖ Business understanding: define what is the purpose of the project; what is desirable to be achieved with this project having in mind that each business has its own unique characteristics
- ❖ Data understanding: study the disposable data in order to be able take the most of it
- ❖ Data preparation: data needs to be cleaned and certain concept specifications must be defined
- ❖ Modelling: decide which model will be applied to the data set
- ❖ Evaluation: evaluates and interprets the discovered results, also determines which next steps should be taken
- ❖ Deployment: the actions that will or will not be taken according to the quality of the achieved results and concerning the purpose of the project

In figure 7 it can be seen CRISP-DM Reference Model with a more detailed approach to the tasks and outputs of this model.



Source: (Shearer, 2000)

Figure 6: Phases of the CRISP-DM Reference Model



Source: (Shearer, 2000)

Figure 7: Tasks and Outputs of CRISP-DM Reference Model

Spiliopoulou follows five steps for the particular problem of improving the success of a web site. The steps are modelled as follows: formal problem specification, data

preparation, data mining, evaluation and interpretation of the results, action upon the mining results. (Spiliopoulou et al., 2001)

2.3.1 Web Mining

Web Mining means mining the World Wide Web, gathering intelligence from using data mining techniques in the web, either in the pages available or in server access logs. Web Mining is typically divided in web usage mining, web structure mining and web content mining. More precisely (Berendt, 2002a), Web usage mining refers to the discovery of user access patterns from Web usage logs. Web structure mining tries to discover useful knowledge from the structure of hyperlinks. Web content mining aims to extract/mine useful information or knowledge from web page contents.

According to (Anand et al., 2004) there are two key applications of web usage mining. They are web measurement and knowledge generation for personalization. Web measurement is concerned with the value created for the business; by using web mining techniques web measurement calculates the success of various marketing efforts and promotions, understanding conversion rates and identifying bottlenecks within the conversion process. Knowledge generation for personalization is all about understanding visitor behaviour in order to serve him better and create guidelines for future interactions. Personalization is a necessary work for the creation of recommender systems. Recommendation systems track past actions of a group of customers to make a recommendation to individual members of the group. (Kim et al., 2002) After knowing your customer you can develop more efficient marketing campaigns for your customers. (Anand et al., 2004)

Define concepts (Srivastava et al., 2000):

- ❖ A *user* is defined as a single individual that is accessing files from one or more Web servers through a browser.

- ❖ A *page view* consists of every file that contributes to the display on a user's browser at one time.
- ❖ A *click-stream* is a sequential series of page view requests.
- ❖ A *user session* is the click-stream of page views for a single user across the entire Web.
- ❖ The set of page-views in a user session for a particular Web site is referred to as a *server session* (also commonly referred to as a *visit*).

Usage Profiles

A very important result of web usage mining is the definition of usage profiles either individual or aggregate. The derivation of accurate and business interesting usage profiles is necessary for the performance of personalization tasks. Good quality profiles should not only reflect users' interests and needs but also be a helpful tool for behaviour definition. Accordingly to Mobasher et al. (2000) aggregate profiles must exhibit three important characteristics: (1) capture possibly overlapping interests of users, (2) provide the capability to distinguish among page views in terms of their significance within the profile, (3) have a uniform representation that allows for the recommendation engine to easily integrate different kinds of profiles.

2.3.2 Web mining tools

“There are many commercial tools which perform analysis on log data collected from Web servers. Most of these tools are based on statistical analysis techniques, while only a few products actually exploit Data Mining techniques.” In (Facca et al., 2005)

When a web user visits a site, his behaviour is registered by the web site server in the form of consecutive URL requests. This information contained in the web site server or,

accordingly to (Spiliopoulou et al., 2001), this log of individual requests must be transformed into a log of sessions, from which navigation patterns should be extracted.

A miner performs the transformation of the web server log into a log of sessions appropriate for mining and the process of navigation pattern discovery (Spiliopoulou et al., 2001), this information is used in web usage mining for different purposes, like personalization, system improvement, site modification, business intelligence or usage characterization. For further details please read chapter 2.4.3 How to generate knowledge from data?, subtitle Using discovered knowledge.

2.4 Data

Data generated by on-line businesses ranges from tens of Megabytes to several hundred Gigabytes per day. Complexity of the web infrastructure and the focus on scalability has led to numerous data quality issues related to page view identification, visitor identification and robot activity filtering. Prior to knowledge being discovered, this data must be cleaned, requiring large processing capabilities. The processed data must then be loaded into an optimised warehouse before even the simplest of statistics can be generated based on the data collected. This in turn means that businesses need to make large investments in hardware and software before they can start gaining the benefits from analysing the data. Analysing the data collected on-line on its own has not provided significant business benefits. (Anand et al., 2004)

In this chapter we will go through the several steps that concern data, from how to collect data to how to generate knowledge from data.

2.4.1 How to collect data?

Data can be collected by explicitly asking the user for the information or implicitly by recording user behaviour (Anand et al., 2004). Each of these forms has its advantages and disadvantages. Collecting data explicitly implies to depend on the willingness of the users to fill forms or rate products/services on the web site. This means that this kind of information is subject to users' bias. User may give incorrect data and even if they are

truly committed in rating a product/service it is always quite subjective. By collecting data implicitly we are analysing user true patterns, if he clicks in a certain link it is probably because he wants to see it, if he spends time reading an article it is probably because he is actually interested in it, if he first downloads one file than another it is probably because he is more interested in the first. In the context of the web, the most commonly available data (which is collected implicitly) is web logs that document visitor navigation through the web site.

2.4.2 What sources should we use to collect data?

Web Usage Mining applications are based on data collected from three main sources (Facca et al., 2005):

- ❖ Web Servers: It will be referred in more detail when discussing data collected from Web Servers (Server data).
- ❖ Proxy servers: Many Internet Service Providers (ISPs) give their customer proxy server services to improve navigation speed through caching. In many respects, collecting navigation data at the proxy level is basically the same as collecting data at the server level. The main difference in this case is that proxy servers collect data of *groups of users* accessing *huge groups* of web servers.
- ❖ Web clients: Usage data can be tracked also on the client side by using JavaScript, Java applets, or even modified browsers. These techniques avoid the problems of users' sessions' identification and the problems caused by caching (like the use of the back button).

In terms of the purpose the data is collected for we have at our disposal three kinds of data: Server data, Marketing Data and Knowledge and Web Meta Data. According to (Buchner et al., 1998):

- ❖ Server data: Is generated by the interactions between the persons browsing an individual site and the web server. The *httpd* process that runs on web servers provides a facility to log information on accesses to the server. That data can be divided into log files and query data.
- ❖ Marketing Data and Knowledge: A second type of information which is stored in modern environments is marketing knowledge. This type of domain expertise obtained internally or externally, has usually been formulated by (human or artificial) marketing experts. It can be in the form of target-directed data collated from cross-fertilised sources, or the output of online analytical web mining activities carried out at an earlier stage.
- ❖ Web Meta Data: The last source is data about the site itself, usually generated dynamically and automatically after a site update. Web Meta Data provides the topology of a site, which includes neighbour pages, leaf nodes and entry points. Meta Data also provides information whether a page has been created statically or dynamically and whether user interaction is required or not.

As it was referred above server data can be divided into log files and query data. Following (Buchner et al., 1998) scheme:

- ❖ Query data to a web server is typically generated on electronic commerce sites. The logged query data must be linked to the access log through cookie data and/or registration information.
- ❖ There are three types of log files, namely:
 - ◆ Server logs: Server logs are either stored in the Common Logfile Format or the more Extended Logfile Format. The available field information from the two files is shown in Table 1, where the gray shaded cells indicate support in the extended format only.

Field	Description
date	Date, time, and timezone of request
client IP	Remote host IP and / or DNS entry
user name	Remote log name of the user
bytes	Bytes transferred (sent and received)
server	Server name, IP address and port
request	URI query and stem
status	http status code returned to the client
service name	Requested service name
time taken	Time taken for transaction to complete
protocol version	Version of used transfer protocol
user agent	Service provider
cookie	Cookie ID
referrer	previous page
...	...

Source: (Buchner et al., 1998)

Figure 8: Server Log File Formats

- ◆ Error logs: Error logs store data of failed requests, such as missing links, authentication failures, or timeout problems.
- ◆ Cookie logs: A cookie is a piece of information sent by a Web server to a Web browser. This information is stored on the user's computer as a text file. The information stored in a cookie log helps to ameliorate the transactionless state of web server interactions, enabling servers to track client access across their hosted web pages. The logged cookie data is customisable, which goes hand in hand with the structure and content of the marketing data. Though it is important to refer that (Facca et al., 2005), even if cookies are used, it is still impossible to reliably identify the exact navigation paths since the use of the back button is not tracked at the server level.

2.4.3 How to generate knowledge from data?

If we go back in history we realize that there has been an attempt to understand what is going on web sites since they were created. Web site owners have always been looking for ways of making their websites more interesting, knowing their clients better and making their web business more profitable. But how has site's performance web metrics evolved throughout the years?

According to (Anand et al., 2004) there have been three waves in the advance of web measurement:

- ❖ The first form of measurement was the embedding of counters on web pages that displayed the number of times that a web page was requested from the server.
- ❖ The second generation of web measurement tools used web server logs to produce more detailed statistics. While useful for site administrators to improve site performance, were not useful from the perspective of gaining insights into user behaviour. These tools did not warehouse the data from which the statistics were produced and nor did they provide systematic filtering and visit/visitor identification techniques so as to provide visitor centric analysis.
- ❖ More recently, the focus of the analytics has shifted to visitor behaviour. Tools used for this type of analysis are often referred to as Web analytics tools. The approach taken by these tools is that a web site can be considered to be successful when the objectives of its owner are satisfied.

Preprocessing

Data preprocessing has an essential role in Web Usage Mining applications. According to (Cooley et al., 1999) we can sum up the main problems and possible solutions for creating a reliable user session file. Meanwhile it is important to refer that the preprocessing phase uses as inputs: Server logs, Site files and Statistics.

There are several problems for creating a reliable user session file, such as:

- ❖ Browser caching: Logs from Web servers do not include any information about the use of the *back* button (Facca et al., 2005).
- ❖ Proxy server caching: Proxy caching causes a single IP address (the one belonging to the Proxy server) to be associated with different users' sessions, so that it becomes impossible to use IP addresses as users identifiers (Facca et al., 2005).

These problems might be overcome by the use of:

- ❖ Heuristics: By using additional information about the Web site structure is still possible to reconstruct a consistent path by means of heuristics in order to face the *back* button problem (Facca et al., 2005).
- ❖ Cookies: The definition of cookie can be found in the chapter 2.4.2 What sources should we use to collect data?
- ❖ Cache busting: It is the practice of preventing browsers from using stored local versions of a page, forcing a new download of a page from the server every time it is viewed.
- ❖ Registration: User registration, i.e. forcing the user to login, has the advantage of being able to collect additional demographic information beyond what is automatically collected in the server log, as well as simplifying the identification of user sessions.

However cookies can be deleted by the user and in what concerns to registration, many users choose not to browse sites that require registration and logins, or provide false information, due to privacy reasons.

According to (Facca et al., 2005) the preprocessing of Web logs is usually complex and time demanding. It comprises four different tasks:

- ❖ The data cleaning: Techniques to eliminate irrelevant items are very important for Web log analysis. The discovered associations or reported statistics are only useful if the data represented in the server log gives an accurate picture of the user accesses to the Web site. Irrelevant items, or files requests that the user did not explicitly asked for should be eliminated (Cooley et al., 1999).
- ❖ The identification and the reconstruction of users' sessions: One important step in the preprocessing task is the identification of unique users. It has already been discussed some of the problems of this identification, in practice two users with the same IP address that use the same browser on the same type of machine can easily be confused as a single user if they are looking at the same set of pages. Conversely, a single user with two different browsers running, or who types in URLs directly without using the link structure can be mistaken for multiple users (Cooley et al., 1999).

Another important step is the session identification. The goal of session identification is to divide the page accesses of each user into individual sessions. The simplest method of achieving this is through a timeout, where if the time between page requests exceeds a certain limit, it is assumed that the user is starting a new session. Cooley et al. (1999) seems to agree with Catledge and Pitkow (1995)¹ who established a timeout of 25.5 minutes based on empirical data.

Another problem in reliably identifying unique user sessions is determining if there are important accesses that are not recorded in the access log. This problem is referred to as path completion. Methods similar to those used for user identification can be used for path completion.

¹ L. Catledge and J. Pitkow. Characterizing browsing behaviors on the World Wide Web. *Computer Networks and ISDN Systems*, 27(6) 1995.

- ❖ The retrieving of information about page content and structure: The vast majority of Web usage mining applications use the visited URLs as the main source of information for mining purposes. URLs are however a poor source of information since; for instance, they do not convey any information about the actual page content. Web content mining and Web structure mining can be of great help to enrich the Web log data (Facca et al., 2005).
- ❖ The data formatting: Once the appropriate preprocessing steps have been applied to the server log, a final preparation module can be used to properly format the sessions or transactions for the type of data mining to be accomplished (Cooley et al., 1999).

But where should we store data for performing the preprocessing phase and where should we save data for extracting knowledge from it?

According to Fayyad (Fayyad et al., 1996a) a data warehouse is where an entity can store its electronically obtained data and also cleaning transactional data to make them available for online analysis and decision support.

It seems of importance to clarify certain concepts in order to proceed. OLAP tools focus on providing multidimensional data analysis which according to Fayyad (Fayyad et al., 1996a) is superior to SQL (a standard data manipulation language) in computing summaries and breakdowns along many dimensions. OLAP (Online Analytical Processing) provides quick answers to analytical queries of multidimensional nature. The output of an OLAP query is typically displayed in a matrix format. OLAP is a popular approach for analysis of data warehouses.

Knowledge discovery techniques/Pattern discovery

Data mining uses several methods and techniques to extract knowledge from large datasets. Some of these methods/techniques that have been applied to the web domain for **pattern discovery** will be shortly referred next according to (Srivastava et al., 2000):

- ❖ Statistical Analysis: Many web traffic analyses produce a periodic report containing statistical information such as the most frequently accessed pages, average view time of a page or average length of a path through a site.
- ❖ Association rule discovery: Association rule generation can be used to relate pages that are most often assessed together in a single server session. In the context of web usage mining.
- ❖ Clustering: Clustering is a technique to group together a set of items having similar characteristics. In the web usage domain, there are two kinds of interesting clusters to be discovered: usage clusters and page clusters. Clustering of users tends to establish groups of users exhibiting similar browsing patterns. Clustering of pages will discover groups of pages having possibly related content, since each group tends to be visited by a community of users.
- ❖ Classification: Classification consists on putting a data item into one of several predefined classes. In the web domain, one is interested in developing a profile of users belonging to a particular class or category.
- ❖ Sequential patterns: The techniques for sequential pattern discovery attempt to find inter-session patterns such that the presence of a set of items is followed by another item in a time-ordered set of session or episodes.

“The motivation behind pattern analysis is to filter out uninteresting rules or patterns from the set found in the pattern discovery phase. The exact analysis methodology is usually governed by the application for which web mining is done. The most common form of pattern analysis consists of a knowledge query mechanism such as SQL. Another method is to load usage data into a data cube to perform OLAP operations.” (Srivastava et al., 2000)

Using discovered knowledge

“Using discovered knowledge: includes incorporating this knowledge into the performance system, taking actions based on the knowledge, or simply documenting it and reporting it to interested parties, as well as checking for and resolving potential conflicts with previously believed (or extracted) knowledge.” (Fayyad et al., 1996a)

The results produced by the mining of Web logs can be used for different purposes. According to (Facca et al., 2005) e (Srivastava et al., 2000) the main application areas for the interesting information which was gathered during the Web usage mining project are the following:

- ❖ **Personalization:** The main purpose of Web Usage Mining is to provide individualized information about the web user. Being able to anticipate the web user behaviour allows the introduction of a recommender system application. A recommenders system's objective is to suggest something that might be of interest to the web user. Personalization systems are usually classified in two dimensions: the individual or the collaborative dimension. Referring to (Anand et al., 2005) when a personalization system builds a individual model of user likes and dislikes and uses this profile to predict future interactions with that user this approach is referred as content-based filtering systems. When a personalization system uses not only the profile for the active user but also other users with similar preferences for item recommendation, this approach is referred as social or collaborative filtering.
- ❖ **System Improvement:** Web usage mining provides an insight on the web traffic behaviour, which in turn allows taking measures concerning the performance of web servers and web based applications. Web usage mining might be used to develop pre-fetching and caching strategies and also to provide patterns which are useful for detecting intrusion, fraud, attempted break-ins, among others.
- ❖ **Site Modification:** Web usage mining techniques can provide guidelines for improving the design of web applications. For a web site to achieve a good

usability level it is necessary to pay close attention to the site content and structure. Web usage mining gives detailed information on user behaviour and its patterns over the web site supplying the designer with the necessary information for taking redesign decisions.

- ❖ **Business Intelligence:** For companies which invest in e-commerce businesses it is of extreme importance to know how their customers are using their web site and why they are getting that feedback from them. The continuity of their business depends on this information. Web usage mining techniques can have an effective advantage on business specific issues as: customer attraction, customer retention, cross sales and customer departure.

Having reached this point of awareness about data and how to use the discovered knowledge extracted from the disposable data, we are left with a question that urges to be answered: How does all this actually is gone help my e-commerce site? Please proceed to next chapter.

2.5 How to measure the success of a site?

“For a business deploying a personalization system, accuracy of the system will be little solace if it does not translate into an increase in quantitative business metrics such as profits or qualitative metrics such as customer loyalty” (Anand et al., 2005)

Spiliopoulou et al. (2001) have made advances in their work mainly in two domains:

- success measures for commercial web sites
- data mining techniques for analyzing web usage

A wide number of success measures have been tried and developed since the first web sites were created. The interest in monitoring the usage of a web site has always existed, right from the beginning.

Berthon et al. (Berthon et al., 1996) modelled the flow of surfer activity on a Web site as a six-stage process; this process has six indexes that measure: the awareness efficiency, the locability/attractability efficiency, the contact efficiency, the conversion efficiency and the retention efficiency. In 2001, Myra et al. (Spiliopoulou et al., 2001) uses two of Berthon's measures of a site's success: the contact efficiency and the conversion efficiency. The first measures how effectively the organization transforms Web site hits into visits, which means the number of users that spent at least a user-defined minimum amount of time exploring the site. The second measures the capability to turn visitors into purchasers, which is measured by the ratio of users that after exploring the site also purchased something. In this study Myra et al. (Spiliopoulou et al., 2001) defined as their goal not only to measure but also to improve a Web site success.

Ruiz Mafé et al. (2006) have made an interesting work on defining an online consumer typology by segmenting consumers according to their web behaviour and web purchases. This study intends to make the web site's marketing actions more profitable and obtain competitive advantages.

3 The Problem

3.1 *CRISP*

The success of a data mining project depends on the development and application of a standard process. CRISP-DM provides a generic basis for such a process that has been tried and tested in industry. (Anand et al., 2004)

As described in chapter 2.3 Data mining, according to Shearer (Shearer, 2000) CRISP-DM (Cross Industry Standard Process for Data Mining) methodology organizes the process of data mining in six big phases:

Business understanding: define what the purpose of the project is; what is desirable to be achieved with this project having in mind that each business has its own unique characteristics

Data understanding: studying the disposable data in order to be able take the most of it

Data preparation: data needs to be cleaned and certain concept specifications must be defined

Modelling: decide which model will be applied to the data set

Evaluation: evaluates and interprets the discovered results, also determines which next steps should be taken

Deployment: the actions that will or will not be taken according to the quality of the achieved results and concerning the purpose of the project

3.2 Business Understanding

When a company owns an electronic commerce website it is interested in knowing if their site has a high number of visitors or how their users value the site's image but they are much more interested in profiting money out of their investment. The focus is purely in making money otherwise there is no point in doing business. If what it takes to get the return on the capital invested is to pay attention to your users' needs that is what will be done. (Sterne, 2002) (Ledford et al., 2007)

The data used in this dissertation analysis was supplied by INTROduxi, a Portuguese company who owns an electronic commerce website (www.introduxi.pt). This website has its core business in selling informatics products to retailers. It is a business to business website. To be a registered user of this website, the client must be a company with an activity concerning informatics selling or assistance. The company has a policy of selling through the on-line channel as well as through the call centre placed at the headquarters (we will call this channel the "classic channel"). As a consequence the website is used not only with the purpose of electronic commerce but also for allowing registered users to check what products exist and if they are available, using afterwards the call centre. Inclusively there are some products that the price is not available, even after being added to the shopping cart. In these situations the user is invited to contact his account manager in the call centre.

The company INTROduxi was constituted in July of 1995. According to the information made available on their website there is evidence of their ability to distinguish in the computer Portuguese market.

1995 – The company INTROduxi initiates its business in July of 1995 with just two people. In this same year they managed to achieve the top 100 biggest companies in the computer business.

1997 – The journal Computerworld, in its edition on the 1st October 1997, considers INTROduxi as the 3rd computer company which has most grown in Portugal.

2002 – INTROduxi is the first computer company in the north of Portugal to have a NP EN ISO 9002 certification.

2003 – In September of 2003 the upcoming realized with the change of facilities to a new warehouse, build from scratch with the company's resources, having a four times bigger area than the old one. At this time INTROduxi had 90 employees, representing the sustained growth of this company.

The warehouse is located 25 meters from the old one. It was built thinking about the customer. It has a modern functional design, with four floors, two over the ground and two beneath it. There is a total area of 2.400 square meters.

In <http://www.introduxi.pt/showdoc.asp?documentid=40>

3.3 Data understanding

There were several meetings with the staff from Introduxi in order to understand what information would be of interest to their business but also to achieve agreement in what data would be disposable for analysis. Introduxi made available the log files of the website and the selling registries.

Tipo de Produto	Type of product
ACESSÓRIOS	ACCESSORIES
COMPONENTES	COMPONENTS
COMPUTADOR	DESKTOPS/LAPTOPS
CONSUMÍVEIS	CONSUMABLES
IMAGEM	VIDEO
PERIFÉRICOS	PERIPHERALS
REDES/COMUNICAÇÃO	NETWORKING
SOFTWARE	SOFTWARE

Table 1: List of products by type.

Introduxi organizes its products into types of products and families of products. Each type of product contains several families. Since the website is Portuguese and the products' names are in Portuguese we made a translation into English for better

understanding. The Table 1 shows the list of products available at the website by type. In Table 2 are listed some products by family. The full list of products by type and family is presented in **Erro! A origem da referência não foi encontrada..**

Família	Family of Product	Product Code Reference
BATERIAS	BATTERIES	BAT
CABOS	CABLES	CB
MICROFONES	MICROFONES	MICROS
COOLERS	COOLERS	COOLER
CPU	CPU	CPU
DISCOS	HARD DRIVES	HD
FONTES	POWER SUPPLIES	PW
MEMORIAS	MEMORIES	RAM
MOTHERBOARDS	MOTHERBOARDS	MBA
PLACAS GRAFICAS	VIDEO CARDS	PLGRA
PLACAS REDE	NETWORK CARDS	PLR
PLACAS SOM	SOUND CARDS	PLSOM
PLACAS VIDEO	TV CARDS AND VIDEO DEVICES	PLVD
GOLDNET	GOLDNET	GOLD
PORTATIL	LAPTOP	PORT
CD	CD	CD
COMPATIVEIS	COMPATIBLES	COMPATIVEL
DVD	DVD	DVD
CAMARAS VIDEO	CAMCODERS	CV
CARTOES MEMORIA	MEMORY CARDS	CARTMEM
DVD SALA	DVD PLAYERS	DVDSALA
DVDPORT	PORTABLE DVD PLAYER	DVDPORT
DVDRW SALA	DVD RECORDERS	DVDRWSALA
BUNDLES	BUNDLES	BUNDLE
DISCOS FLASH USB	USB FLASH DRIVE	FLASHU
JOYSTICKS	JOYSTICKS	JOYSTICKS
RATOS	MOUSE	RATO
UPS	UPS	UPS
WIRELESS	WIRELESS	WIRELESS
ANTIVIRUS	ANTIVIRUS	ANTV
SOFTWARE	SOFTWARE	SWS

Table 2: List of products by family.

3.4 Data preparation

The data used in this dissesation went through a pre-processing phase performed by the “Laboratório de Inteligência Artificial e Apoio à Decisão” (LIADD).

The pre-processed data was stored in a data warehouse and the necessary queries were performed using SQL language.

4 Working on the Problem

4.1 Modelling

The pre-processed data was stored in a data warehouse and the necessary queries were performed using SQL language.

Spiliopoulou et al. (2001) define ratios for measuring and improving the site's success:

- the contact efficiency of an action page A
- the relative contact efficiency of an action page A
- the conversion efficiency of a page P for a target page T

In order to understand these measures it is necessary to know the definition of variables used in their calculation. The following concepts are defined according to Spiliopoulou et al. (2001).

Action page $\rightarrow$ is a page whose invocation indicates that the user is pursuing the site's goal

Session $\rightarrow$ is a sequence of consecutive URL requests performed by the same visitor; the boundaries of a session can be specified either by duration or by content

Active session $\rightarrow$ is a session containing at least one activity towards fulfilling the site's goal; according to the definition of action pages, active sessions are those containing an access to at least one action page

Customer session $\rightarrow$ is a session in which the user has achieved the site's goal; the same physical person may behave once as a costumer and once as a non-customer

Contact efficiency of an action page A $\rightarrow$ the ratio of sessions containing this page (A) to all sessions in the log

$$contacteff(A) = \frac{\#(\{S \in Sessions \mid A \in S\})}{\#(Sessions)}$$

Relative contact efficiency of an action page A → the ratio of sessions containing this page (A) to the cardinality of the multiset of active sessions, called aSessions

$$Rcontacteff(A) = \frac{\#(\{S \in Sessions \mid A \in S\})}{\#(aSessions)}$$

Target page → is a page whose invocation indicates that the user has achieved the site's goal

Conversion efficiency of a page P for a target page T → this measure estimates the success of an arbitrary page in helping/guiding the users towards a target page; the ratio of active sessions containing the page P and T (the access to page T is after the access to page P) to active sessions containing P

$$conveff(P, T, G) = \frac{\#(G)}{\#(\{S \in Sessions \mid P \in S\})}$$

4.2 Exploiting Introdusi's data

The period of time it was supposed to be analysed was between the 1st November 2007 and the 31st January 2008. Unfortunately the only month that had complete data was the month of January 2008. All the others had missing days. It was decided that in order to compare the site's navigation patterns and selling data it would be feasible to just consider the month of January 2008.

The homepage of the website has a login area on the top-left corner and a list of the families of products available right underneath. On the top of the homepage the user might navigate directly to Goldnet (the computers' trademark owned by the company),

laptops (*portáteis*), video security systems (*sistemas de videovigilância*) and other options concerning on-line services provided by the company to registered users. Below this banner there is the possibility of finding products through the research option. On the right side of the homepage there is a list of the trademarks that the company represents (including their own trademark: Goldnet) that redirect to the products of that trademark.



Figure 9: Homepage of Introduxi's website

Only registered users have access to prices, Goldnet products and information about the product quantities available. Also important to notice that only registered users can buy the site's products.

The action page is defined as a page whose invocation indicates that the user is pursuing the site's goal. In this way we consider that an action page, in Introduxi's website, would be a page containing information about a specific product. In e-commerce sites there is a pattern that has to be followed to complete the buying process. The action page is part of that process.

Clicking on an action page we will have the following link:

❖ [http://www.introduxi.pt/revendedor/stocks/produto_info.asp?id=\(...\)](http://www.introduxi.pt/revendedor/stocks/produto_info.asp?id=(...))

The target page is defined as a page whose invocation indicates that the user has achieved the site's goal. In this way we consider that a target page, in Introduxi's website, would be a page where the order is confirmed and concluded. Only here the buying process is completed.

Clicking on an action page we will have the following link:

❖ <http://www.introduxi.pt/revendedor/loja/encomendar.asp>

To set the user to buy the products of the website is the ultimate goal of the site's owner.

For the calculation of the different measures it was considered every click on the action page independently of the fact that the user is registered or not. The reason is that some registered users only login when they actually find what they are looking for and want to buy.

At the beginning of this study it has been defined a limit time for each session of 26 minutes. This time was determined having in consideration Catledge and Pitkow (1995)² who established a timeout of 25.5 minutes based on empirical data, please for more detailed information see chapter 2.4.3 How to generate knowledge from data?, subtitle Preprocessing. After talking to the company's technician it was understood that there was no real need for limiting sessions. When a registered user enters the site if he stops navigating for 5 minutes, the session is closed by the web server and the user must login again.

² L. Catledge and J. Pitkow. Characterizing browsing behaviors on the World Wide Web. *Computer Networks and ISDN Systems*, 27(6) 1995.

4.2.1 The buying process

The buying process is the process that leads the user to reach the site's goal by confirming and concluding the order. The buying process, in Introduxi's website, needs 5 steps to be completed. Every link of the process begins in the same way (<http://www.introduxi.pt/revendedor>). What comes next is what characterizes which step is each:

1. **.../stocks/produtos.asp?tipo=(...)&description=(...)&objRef=(...)**: In this step the user chooses which kind of product he is interested in
2. **.../stocks/produto_info.asp?id=(...)**: In this step the user selects the specific product he has chosen
3. **.../loja/comprar.asp?id=(...)**: In this step the user adds the product to his shopping cart
4. **.../loja/encomenda.asp**: In this step the user gives his accordance to conclude the order
5. **.../loja/encomendar.asp**: In this step the user has his last opportunity to check all the items in the shopping cart before definitively concluding his order

We have to make some notes about the buying process of the Introduxi's website because these particularities will have influence in the calculation and interpretation of the measures.

- ❖ It is important to notice that it is possible to jump over step 2, when adding a specific product to the shopping cart. After selecting the kind of product the user is interested in (step 1), instead of having a closer look to the specific product he has chosen (step 2) he immediately adds it to the shopping cart (step 3).

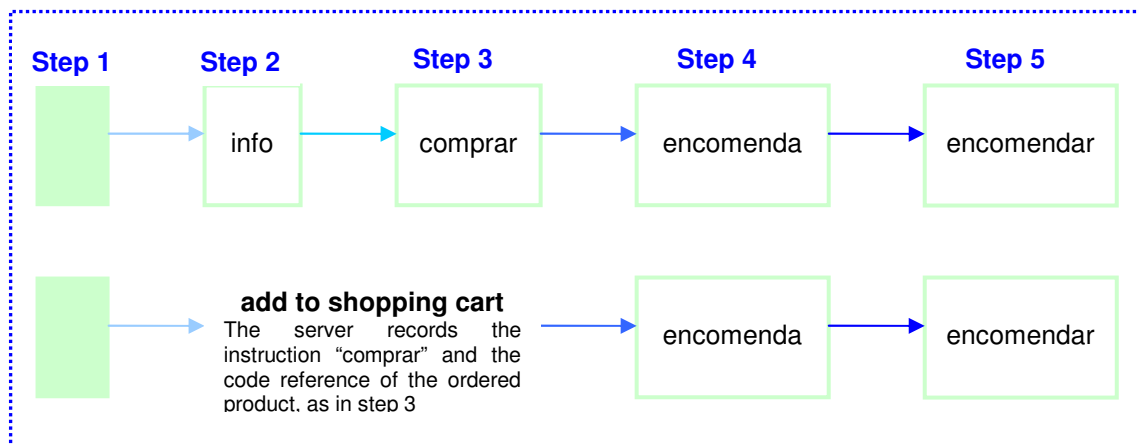


Table 3: The two different ways of completing a buying process in the Introduxi's website

- ❖ Usually sessions contain several specific products, but only one conclusion of the order. As a consequence some specific products might be considered part of the order when they were simply visualized (Step 2). For more accuracy we considered only the sessions with products that have been added to the shopping cart (Step3). If these specific products are put out of the basket right before the conclusion of the order there is no way to know. It was assumed that once a specific product was put inside the shopping cart it would be bought.

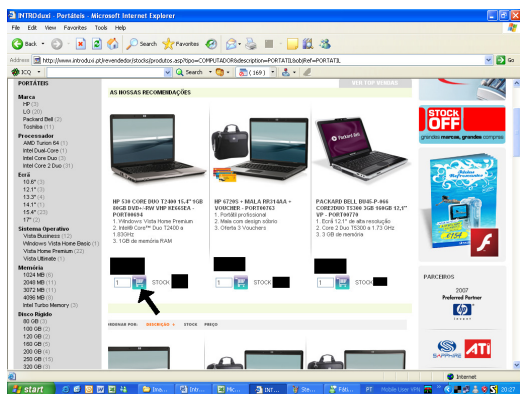


Figure 10: Step 1 of the buying process.

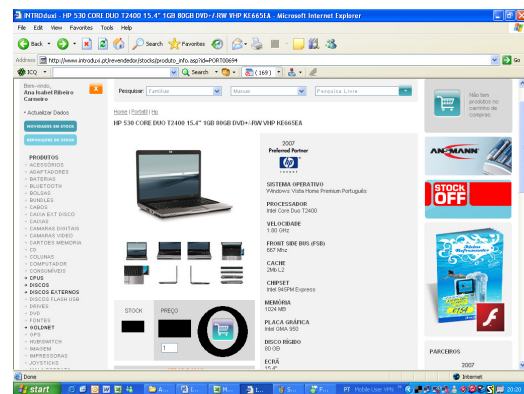


Figure 11: Step 2 of the buying process.

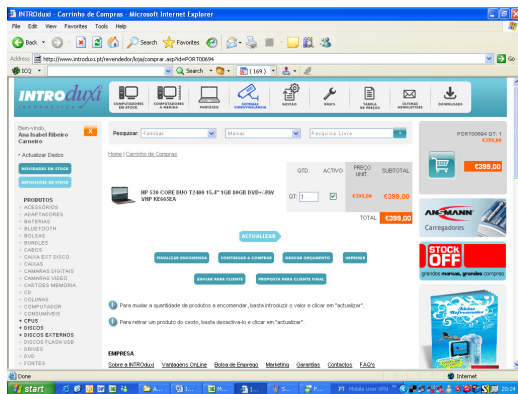


Figure 12: Step 3 of the buying process.

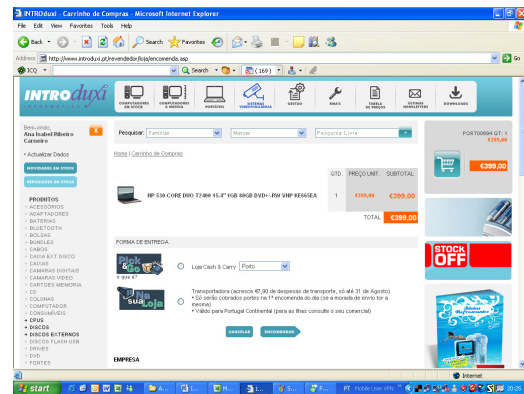


Figure 13: Step 4 of the buying process.

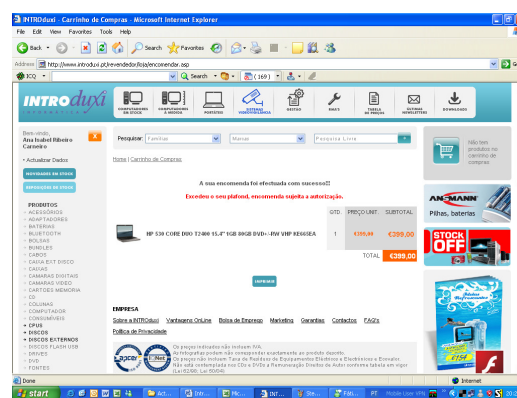


Figure 14: Step 5 of the buying process.

4.2.2 Analysis of selling data

We will now analyse Introduxi's data by first exploring the selling registries. The selling registries or accounting results were analysed in two different perspectives, in terms of turnover per product and in terms of number of products ordered. Finally we made a confrontation between the products with the best turnover and the products with the highest number of quantities ordered. This allowed us to get better acquainted with the data.

Before proceeding we will define turnover as the amount of sales over a specified period. When we refer to the turnover rate of a certain product we mean the ratio of a certain product sales to the total amount of sales (including all products traded).

Analysing turnover

Analysing the selling data we can conclude that 33% of the type of products (Componentes, Computador e Periféricos) are responsible for the major slice of the turnover both through the net channel (around 76%) and through the “classic” channel (around 75%).

The main difference between the two channels is that the computers’ slice is even more representative on the classic channel than it is on the net channel.

The services slice aggregates the following reference codes:

- NSCLUB – refers to offers from the Introduxi’s club, consists in an exchange of accumulated points to products
- NSEXPOSITOR08 – refers to marketing material given to customer
- RMA – is concerned with the delivery of RMAs
- NSPORTES – refers to the costs of products’ delivery, these costs are paid by the customer; this code is the most recurrent in the services slice

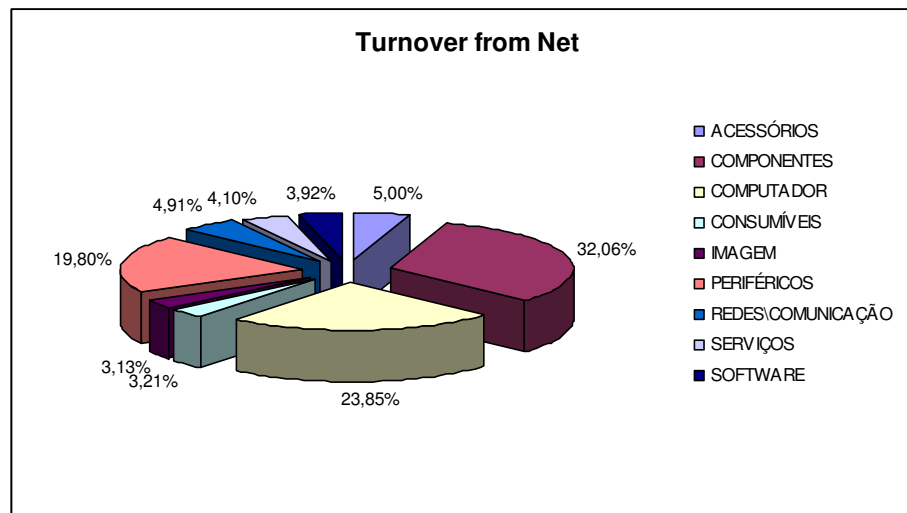


Chart 1: How the turnover from the Net is split into types of products.

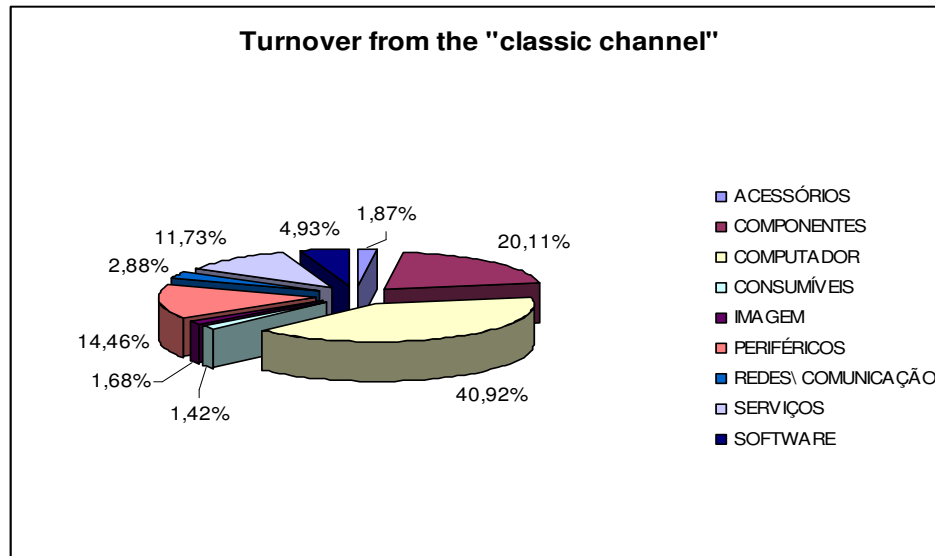


Chart 2: How the turnover from the “classic channel” is split into types of products.

A closer look inside these slices shows that there are 5 families of products that just by themselves explain around 40% of the turnover through the net channel and around 61% of the turnover on the classic channel. These five families represent around 5% of the all families. To be having the turnover so much dependent on such a small percentage of products might not be a good option. However a month is not representative for this kind of conclusions.

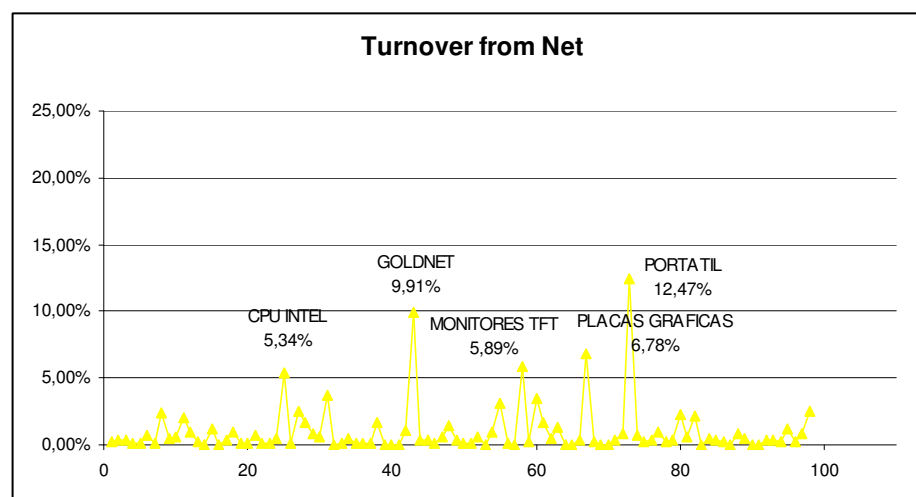


Chart 3: How the turnover from the Net is split into families of products.

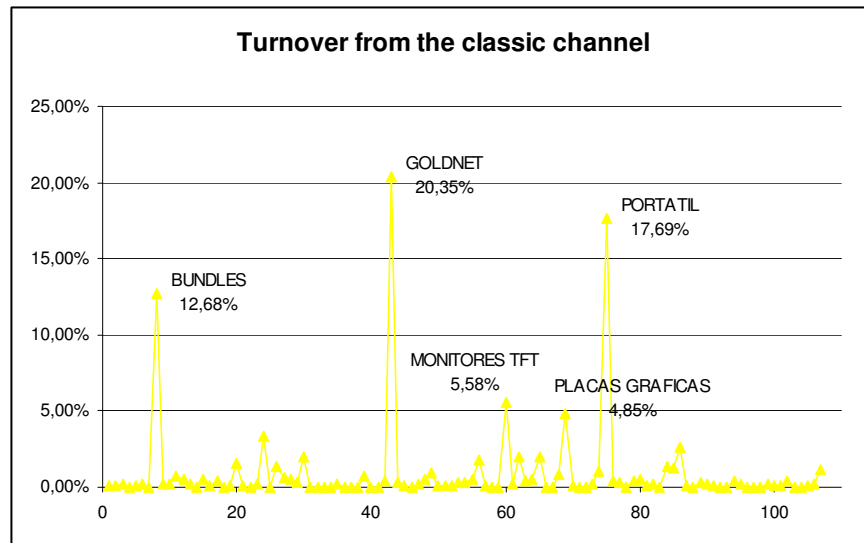


Chart 4: How the turnover from the “classic channel” is split into families of products.

Analysing quantities

The types of products which account for most of the sales are: *Acessórios* (16%), *Componentes* (25%), *Consumíveis* (19%) and *Periféricos* (25%). It is interesting to notice that *Computador*, with a turnover of 24% is responsible for only 1% of the quantities sold. Of course selling computers and mice is not the same thing. They have different prices and the turnover reflects the quantity $\times$ price. It would be better to confront Net results, but we did not have access to this information.

We should also to refer that *Acessorios* and *Componentes* show a turnover of 5% and 3%, respectively.

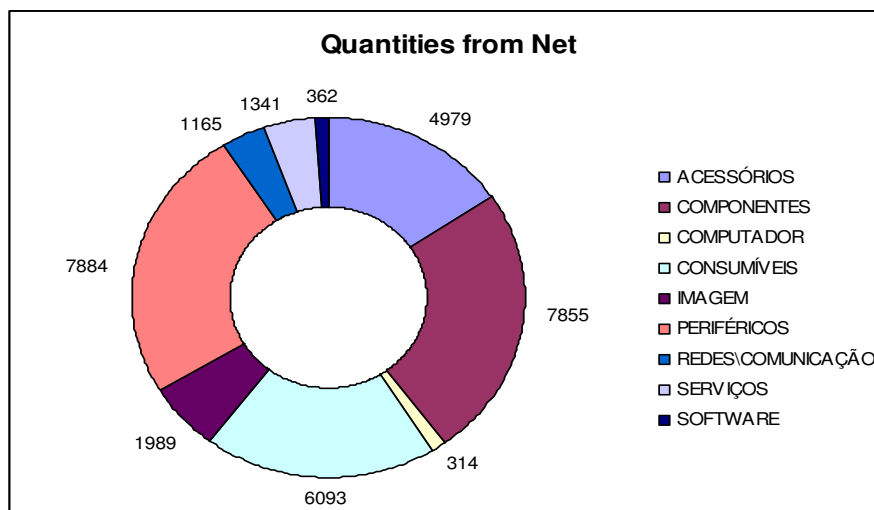


Chart 5: How the quantities are split into types of products.

The average price of the top sales products are the lowest of all. The average price of products sold by Introduxi is 84,07€. The top sales products have their average prices between 4,55€ and 20,99€. The highest average price belongs to laptops (710,36€) which represent 0,35% of the total quantities sold.

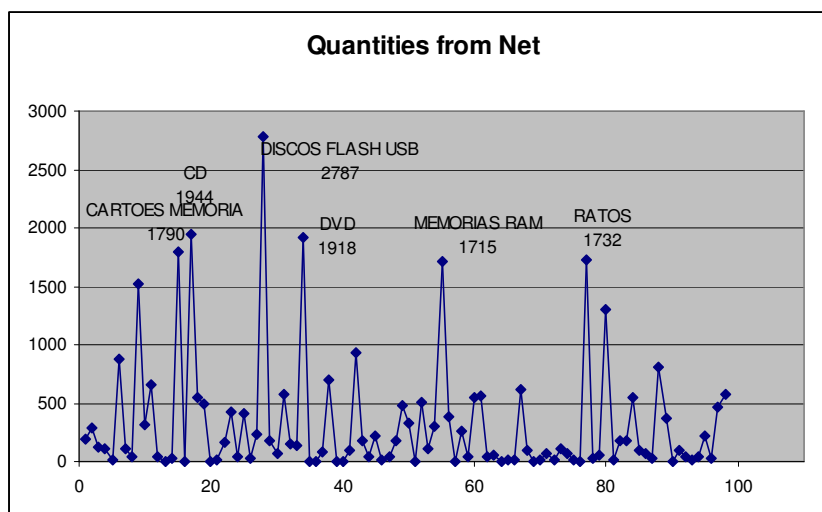


Chart 6: How the quantities are split into families of products.

Turnover versus Quantities

We have analyzed the weight of each family of products in terms of quantities and turnover. Looking at Chart 7 it seems quite reasonable to conclude that the products with the highest weight in quantities are systematically the ones with the lowest average prices and the products with the smallest amount of unities sold are the ones with the utmost average price.

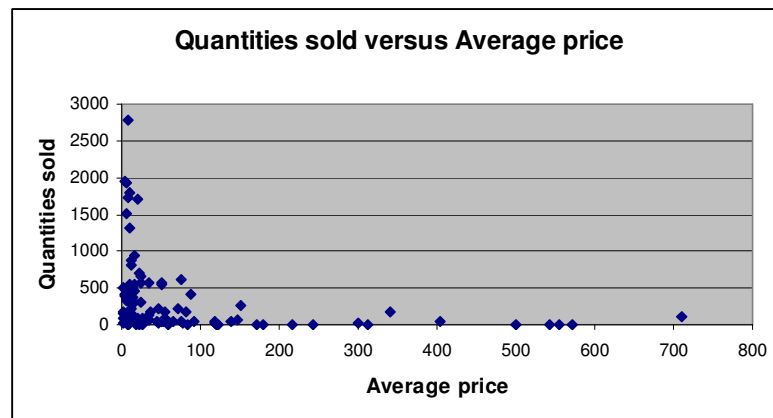


Chart 7: Comparison between the number of quantities sold and the average price by family of product.

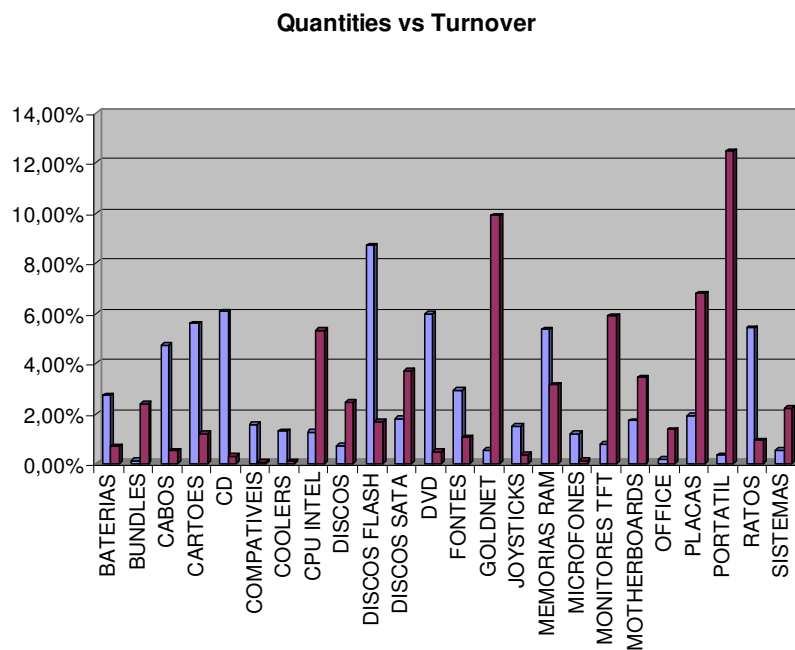


Chart 8: Confrontation between the weight of quantities (%) and the weight of turnover (%) for families of products.

Analysing Chart 8 we find which families of products exhibit the highest discrepancy in terms of quantities sold and turnover. The blue bars are the quantities in percentage and the pink bars are the turnover in percentage.

4.2.3 Log analysis

The analysis of the log file revealed a meaning harmony with the accountancy results already explored.

The number of page requests exhibits a relatively balanced up and down cycle throughout the observation period.

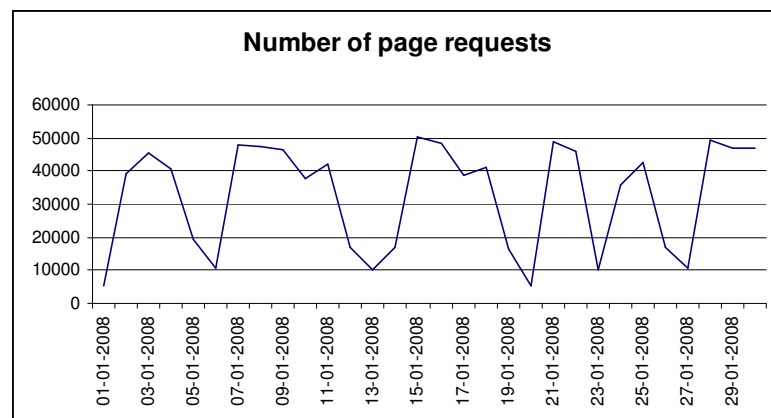


Chart 9: Page request distribution over the observation period

During the observation period there were a total number of 980.945 page requests and a total number of 74.465 sessions. It was also identified 47.590 active sessions.

The active log consisted of 2.302 customer sessions and 45.288 non-customer sessions.

The number of sessions chart resembles quite closely the number of page requests chart. The difference lies in the number of sessions being much smaller than the number of page requests.

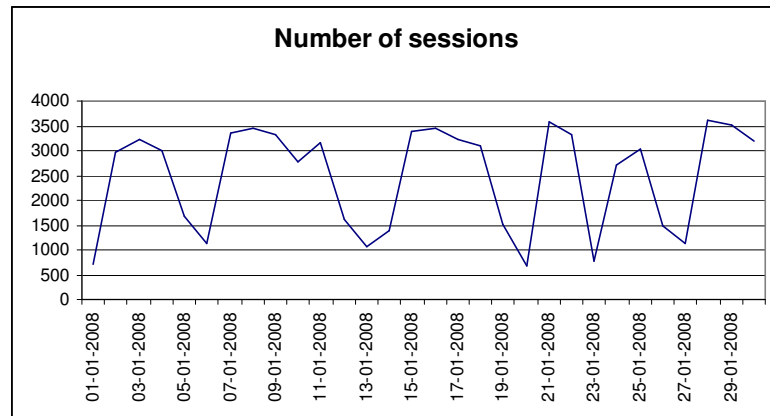


Chart 10: Number of sessions over the observation period

The number of page requests over the week chart illustrates that the site behaves much like in resemblance a 9 to 5 office. There is a peak on Tuesday and another on Friday, although slightly minor. From Tuesday forward there are a descending number of page requests with the lowest happening on the weekend.

The number of orders has a similar behaviour, having its top on Wednesday. To notice though that there were 142 orders on Saturday and 67 on Sunday, quite reasonable for an electronic commerce website designed for retailers' shops.

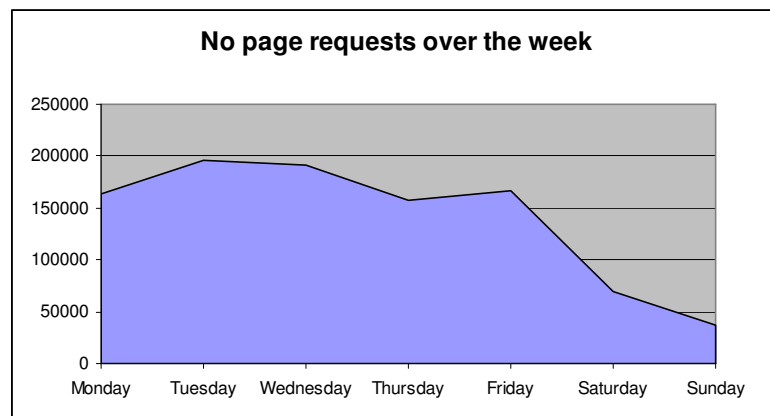


Chart 11: Page request distribution over the week for the observation period.

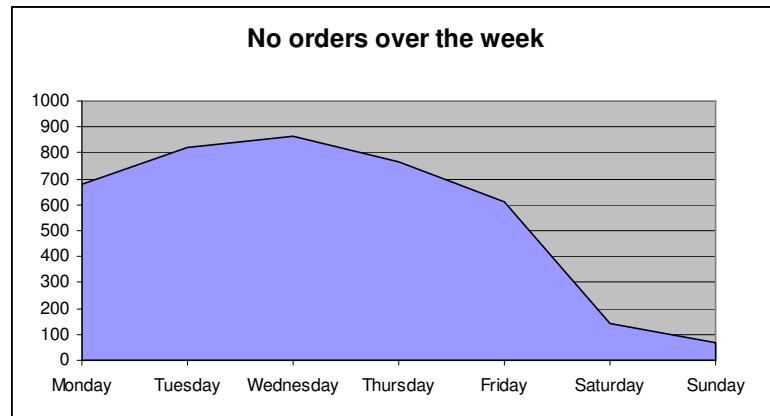


Chart 12: Orders distribution over the week for the observation period.

Another curious thing is the distribution of the same referred parameters during the day. The first peak of orders and page requests happens between 10 and 12 a.m. The second peak happens systematically between 3 p.m. and 6 p.m. for page requests and evidences a confidently number of orders at 6 p.m.

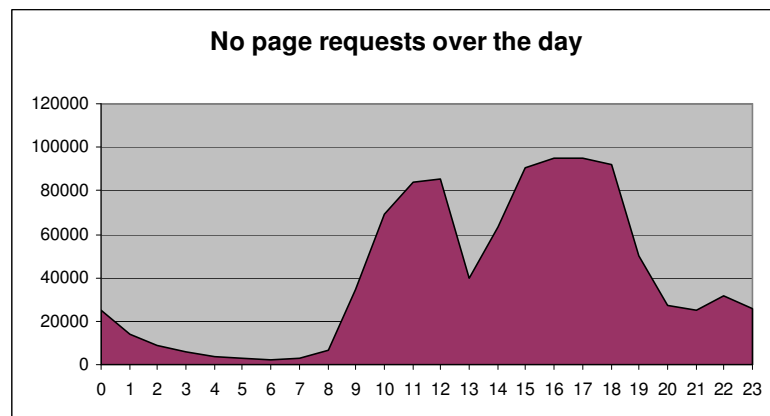


Chart 13: Page request distribution over the day for the observation period

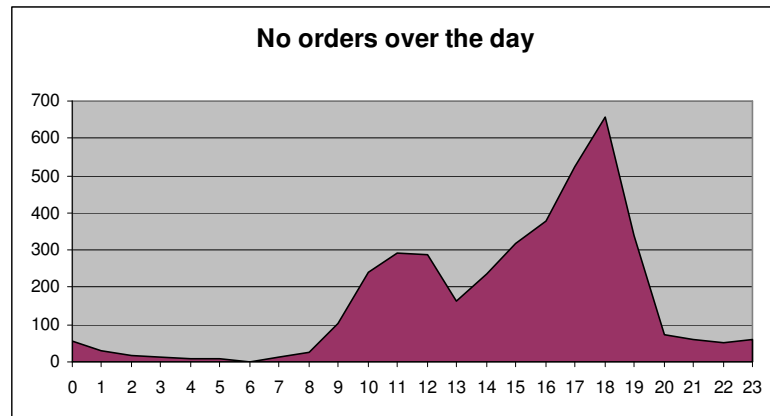


Chart 14: Orders distribution over the day for the observation period

Confronting the top ten clickers and the top ten buyers makes us realize that they do not match. There is not only one repeated name in the two top ten lists. Mr. Sérgio Martinho, the number 4 clicker, is not even a presence in the buyers' table, not only a single order. However he has bought some merchandising on the "classic channel".

The top clicker is an IP address and it is a robot.

And what about Mr. "I'm not a registered user" but "I'm always navigating on the website"! Who does the IP address 74.220.207.87 belong to? This IP is located in the United States, in the city of Orem. It accessed 1.847 different pages in just two days of the month of January, in 4 different sessions. There is no referrer data on this IP address. It seems like another robot.

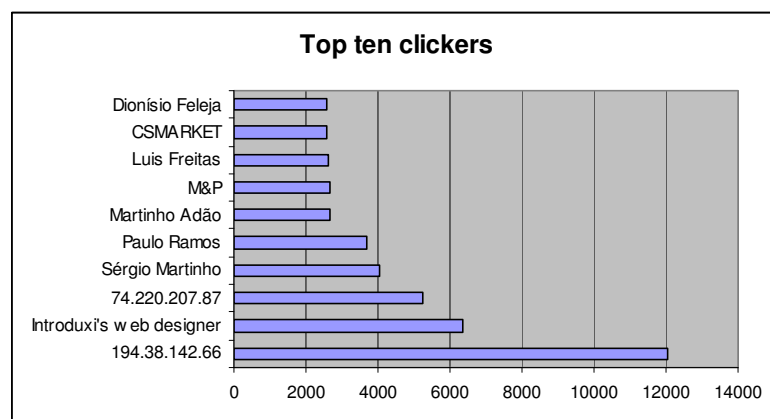


Chart 15: Top ten clickers.

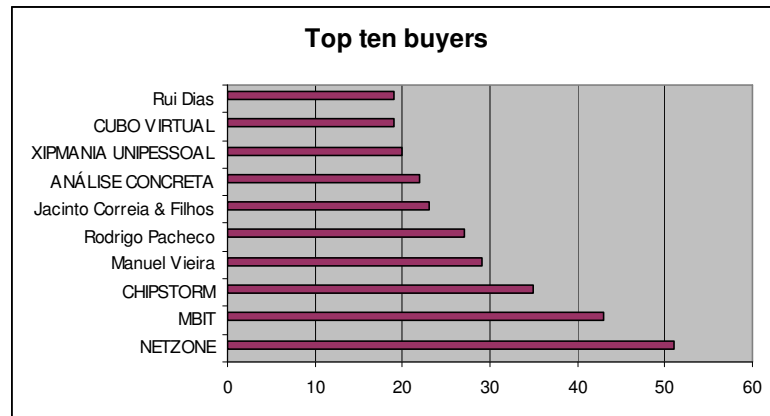


Chart 16: Top ten buyers.

The calculation of the measures presented early in chapter 4.1 Modelling, will permit us to retain certain conclusions on the website's success.

It seems of great importance to explain each of the measures which will be used to study the website's success. According to Spiliopoulou et al. (2001).:

Contact efficiency → the contact efficiency of A is the percentage of sessions in which an attempt to reach the site's goal was made using action page A. By computing this value for each action page, we can (i) identify the impact of each page on the overall success of a site in engaging visitors and (ii) detect pages with low contact efficiency

Relative Contact efficiency → this measure expresses the relative importance of each action page within a site and is appropriate for sites with many action pages and/or a large number of inactive sessions

Conversion efficiency → this measure estimates the success of an arbitrary page in helping /guiding the users towards a target page. With this measure, we can study the impact of each page in the success of the site and identify pages that have low conversion efficiency and require improvement.

To compute the contact efficiency we used the following query:

(Query no 1)

```
SELECT COUNT (DISTINCT session.session_id)
FROM session, usage, page, date
WHERE session.dimension_key = usage.session AND page.dimension_key = usage.page AND
date.dimension_key = usage.date
AND date.date_id BETWEEN '2008-01-01' and '2008-01-31'
AND page.page_id like '%PORT%'
AND page.page_id LIKE '%info%'
```

Note: For further detail on this query please check on appendix 1.

The query will search for all the pages that contain the word “PORT”, the code reference for laptop, as well as the word “info” which appears every time information on a specific product is asked (step 2 of the buying process). Then the query counts all the different sessions respecting this condition. This number is divided by 74.465, the total number of sessions in the log.

$$Contact_Efficiency_{product_x} = \frac{Query1_{product_x}}{No_sessions} = \frac{Query1_{product_x}}{74.465} \quad (Eq. 1)$$

A contact efficiency result of 9,21%, as can be seen in Table 4, means that the impact of the type of product Acessórios on the overall success of the site in engaging visitors is around 9%.

The overall results for the contact efficiency measure are very low. The highest result is for the product laptops, with a value of 4,35%. This does not mean that the site is a failure; it has to be taken in consideration that there are 81 action pages. Meaning a user is not likely to visit such a number of different products every session.

To calculate the Relative Contact Efficiency we used the same query presented above for determining the numerator and the total number of active sessions in the log (47.590) for the denominator. The results were also very poor. Again with the laptop

getting the highest score, 6,80%, meaning that in almost 7% of the active sessions there is a consult to a laptop.

$$\text{Relative_Contact_Efficiency}_{product_x} = \frac{\text{Query1}_{product_x}}{\text{No_active_sessions}} = \frac{\text{Query1}_{product_x}}{47.590}$$

(Eq. 2)

A relative contact efficiency result of 14,41%, as can be seen in Table 4, means that the relative importance of the action page of the type of product Acessórios within the site is around 14%.

To estimate the Conversion Efficiency we used the following query:

(Query no 2)

```
SELECT session.id
FROM session, usage, page, date
WHERE session.dimension_key = usage.session AND page.dimension_key = usage.page AND
date.dimension_key = usage.date
AND date.date_id BETWEEN '2008-01-01' and '2008-01-31'
AND page.page_id like '%PORT%'
AND page.page_id like '%info%'
INTERSECT
SELECT session.id
FROM session, usage, page, date
WHERE session.dimension_key = usage.session AND page.dimension_key = usage.page AND
date.dimension_key = usage.date
AND date.date_id BETWEEN '2008-01-01' and '2008-01-31'
AND page.page_id like '%PORT%'
AND page.page_id like '%comprar%'
INTERSECT
SELECT session.id
FROM session, usage, page, date
WHERE session.dimension_key = usage.session AND page.dimension_key = usage.page AND
date.dimension_key = usage.date
AND date.date_id BETWEEN '2008-01-01' and '2008-01-31'
AND page.page_id like '%encomendar%'
```

Note: For further detail on this query please check on appendix 1.

The first part of the query will search for all the pages that contain the word “PORT”, the code reference for laptop, as well as the word “info” which appears every time information on a specific product is asked (step 2 of the buying process). The second part of the query obligates the link to contain both the code reference of the product and the word “comprar”, which means that the product has been added to the shopping cart (step 3 of the buying process). The third part of the query demands that the session has a page that achieves the site’s goal by finishing the order (step 5 of the buying process). Then the intersect makes that the sessions respecting all the demands are the same. This query was used to determine the numerator of the measure, since the data for the denominator is the same used for the numerator of the Contact Efficiency measure, determined using query 1.

$$Conversion_Efficiency_{product_x} = \frac{Query2_{product_x}}{Query1_{product_x}} \quad (Eq. 3)$$

A conversion efficiency result of 5,10%, as can be seen in Table 4, means that the success of the type of product Acessórios in helping/guiding the users towards a target page is around 5%.

The results for the Conversion Efficiency measure overall are worse than those obtained for the Relative Contact Efficiency measure, exception made to Consumiveis and Software. Notwithstanding the products that perform better are the ones with the lowest average prices. For instance, the mouse pads (type of product Acessorios) performing a conversion efficiency of 11,11% cost in average 2,68€. On the other side products like laptops, computers or LCD with relative high results on the Relative Contact Efficiency measure (6,80%, 2,64% and 4,79%, respectively) have bad marks on the conversion efficiency (0,83%, 1,11% and 2,28%, respectively) and high average prices (710,36€, 341,42€ and 555,07€, respectively).

When analysing the global results of the measures we reach the same conclusions. The types of products that are most popular (Contact efficiency and Relative Contact

Efficiency) are the ones with the highest turnover (Componentes and Perifericos). The types of products that sell (Conversion Efficiency) are the ones with good results on the quantities chart (Acessorios, Componentes, Consumiveis, Perifericos).

Product type	Contact efficiency	Relative Contact efficiency	Conversion efficiency
ACESSORIOS	9,21%	14,41%	5,10%
COMPONENTES	20,97%	32,82%	4,56%
COMPUTADOR	8,26%	12,92%	0,99%
CONSUMIVEIS	2,26%	3,53%	4,52%
IMAGEM	2,55%	3,99%	3,11%
PERIFERICOS	16,83%	26,33%	4,36%
REDES\COMUNICAÇÃO	6,57%	10,28%	3,58%
SOFTWARE	1,35%	2,11%	3,89%

Table 4: Efficiency results per product type considering the entire log

A positive surprise is the type of product Software, as can be seen in Table 4, with low results on the measures Contact efficiency and Relative contact efficiency, turns up with a good result on the conversion efficiency. Another surprise but for its negative impact is the type of product Computador, with a not so good value for conversion efficiency measure: 0,99%.

4.2.4 Log analysis - Customer Session

In Spiliopoulou et al. (2001) the measures: Contact efficiency, Relative Contact efficiency and Conversion efficiency are calculated considering the customer log. The customer log contains a certain number of “customer sessions” and a certain number of “non-costumer sessions”. As it has been referred earlier, and accordingly to Spiliopoulou et al. (2001), a customer session is a session in which the user has achieved the site’s goal; an active session is a session containing at least one activity towards fulfilling the site’s goal. Therefore, and considering that a target page is only reachable via some action page, a customer session is always an active session. Similarly to active sessions and inactive sessions, a session can be uniquely characterized as a customer session from the pages it contains.

The assumption that a target page is only reachable via some action page is not completely true for the case of Introdusi, even though this assumption has been considered for all calculations. For more details on this subject please see Chapter 4.2.1 The buying process.

To determine all customer sessions and considering that a target page is only reachable via some action page, it is necessary to assure that a session contains an action page (by requiring the word 'info' to be present in one of the sessions' link) as well as a target page (by requiring that the word 'encomendar' to be present in one of the sessions' link). To calculate the number of customer sessions we have used the following query:

(Query no 3)

```
SELECT DISTINCT session.id
FROM session, usage, page, date
WHERE session.dimension_key = usage.session AND page.dimension_key = usage.page AND
date.dimension_key = usage.date
AND date.date_id BETWEEN '2008-01-01' and '2008-01-31'
AND page.page_id like '%info%'
INTERSECT
SELECT DISTINCT session.id
FROM session, usage, page, date
WHERE session.dimension_key = usage.session AND page.dimension_key = usage.page AND
date.dimension_key = usage.date
AND date.date_id BETWEEN '2008-01-01' and '2008-01-31'
AND page.page_id like '%encomendar%'
```

Note: For further detail on this query please check on appendix 1.

The measure Relative contact efficiency was computed using the result of the previous query as a denominator (2.302 costumer sessions). The numerator of this measure was calculated using the following query:

(Query no 4)

```
SELECT session.id
FROM session, usage, page, date
WHERE session.dimension_key = usage.session AND page.dimension_key = usage.page AND
date.dimension_key = usage.date
AND date.date_id BETWEEN '2008-01-01' and '2008-01-31'
AND page.page_id like '%PORT%'
AND page.page_id like '%info%'
INTERSECT
SELECT session.id
FROM session, usage, page, date
WHERE session.dimension_key = usage.session AND page.dimension_key = usage.page AND
date.dimension_key = usage.date
AND date.date_id BETWEEN '2008-01-01' and '2008-01-31'
AND page.page_id like '%encomendar%'
```

Note: For further detail on this query please check on appendix 1.

The first part of the query obligates the link to contain both the code reference of the product and the word “info”, which means that the link is an action page. The second part of the query demands that the session has a page that achieves the site’s goal by finishing the order. Then the intersect makes that the sessions respecting both demands are the same.

$$\text{Relative_Contact_Efficiency}_{\text{product}_x} = \frac{\text{Query4}_{\text{product}_x}}{\text{No_customer_sessions}} = \frac{\text{Query4}_{\text{product}_x}}{2.302}$$

(Eq. 4)

A relative contact efficiency result of 28,97%, as can be seen in Table 5, means that the relative importance of the action page of the type of product Acessórios within the site, considering only customer sessions, is around 29%.

The conversion efficiency is determined using, as the numerator, the query 2 presented in the “Log Analysis” chapter. The data for the denominator is the same used for the

numerator of the relative contact efficiency measure (considering customer session log) and is determined using query 4.

$$Conversion_Efficiency_{product_x} = \frac{Query2_{product_x}}{Query4_{product_x}} \quad (Eq. 5)$$

A conversion efficiency result of 52,47%, as can be seen in Table 5, means that the success of the type of product Acessórios in helping/guiding the users towards a target page, considering only customer sessions, is around 52%.

It is the time to find out if the results for the efficiency measures obtained from the customer session log are very different from the results of the measures achieved from the entire log.

A first analysis shows that the measures reach higher values using the customer session log. This means that the relative importance of each action page within the site (Relative Contact efficiency) increases among the buying customers; also the success of the site (Conversion efficiency) is higher among the buying customers. These seems like two very straightforward sentences. But we should take a closer look:

→ The measure of Relative Contact efficiency showed a decrease to one type of product (Computador) when compared to the same result considering the entire log. The action pages grouped in the type of product Computador are less important among the sessions of the buying customers.

→ The most popular types of products for the Relative Contact efficiency results in the previous chapter (Componentes and Periféricos) are still the most popular when considering the customer session log.

→ When comparing the results for the measure Conversion efficiency (entire log vs customer session log) we observe that for customer sessions the results are on average around 50%. This means that, on average, each action page has a 50% success on

leading the customer to the target page. This is a pretty good result for the conversion efficiency measure within the buying customers.

Product type	Relative Contact efficiency	Conversion efficiency
ACESSORIOS	28,97%	52,47%
COMPONENTES	54,47%	56,78%
COMPUTADOR	9,30%	28,50%
CONSUMIVEIS	5,00%	66,09%
IMAGEM	5,13%	50,00%
PERIFERICOS	43,83%	54,11%
REDES\COMUNICAÇÃO	15,68%	48,48%
SOFTWARE	3,39%	50,00%

Table 5: Efficiency results per product type considering customer session

→ If we sort (by importance) the conversion efficiency results of both analysis (entire log vs customer session log) it is interesting to notice that almost all the products changed their order of importance, with the exception of the type of product Computador that maintained the last position in both the analysis. The most important products in terms of conversion efficiency results, despite their order, are still the same four (Acessorios, Componentes, Consumiveis and Perifericos) in the customer session log analysis.

5 Evaluation and Deployment Plan

5.1 Business Intelligence Analysis

“... Adapting to Internet user profiles and needs and identifying the segments of consumers who spend more online are essential for the efficient use of resources invested in Internet marketing campaigns and to help companies obtain a competitive advantage.” (Ruiz Mafé et al., 2006)

Not all customers are equally important to the success of a company. There are customers more important than others when it comes to making money. Different customers deserve different management politics.

Quantities ordered versus Customer share

We have analyzed the “Top ten buyers” in terms of orders that have been made. This data was collected from the log. But are these customers really the ones responsible for the success of the company? Having a high number of orders does not necessarily mean buying a lot (in terms of value).

Name	Share of customer	Quantity	Average price
COMPOSWARE, SC	0,08%	611	7,18 €
MBIT, LDA	1,27%	471	36,06 €
Manuel Henrique Vieira	1,69%	434	43,43 €
Alt4-Informatica	0,56%	387	29,62 €
MPORTO-Manuel Alvarino R.Porto	0,55%	353	28,47 €
Carla Dolores Vieira Pestana Duarte	0,53%	321	31,41 €
NETZONE, LDA	1,02%	319	47,29 €
JCTEK - Inf. Prest. Serv. Equip. Consumo Unipessoal Lda	0,17%	312	14,45 €
EXTREMIS-COM. DE PRODUTOS INFORMATICOS	0,23%	299	19,29 €
Nuno Correia Informática Unipessoal, Lda	0,48%	295	27,61 €

Table 6: Ten customers that bought more quantities of products regardless of the product.

The selling data has something interesting to reveal. The ten customers who buy the highest quantities are not the same as the ten customers with the highest customer share. The exception is Manuel Henrique Vieira, MBIT, Lda and NETZONE, Lda. Another thing that draws our attention is the average price. High quantities most of the times mean very low average price by product. The opposite is also true. The customer VGER only bought 26 products but with an average price of 177,08€.

Name	Share of customer	Quantity	Average price
Manuel Henrique Vieira	1,69%	434	43,43 €
LUDIÁTICA SISTEMAS INFORMÁTICOS LDA	1,33%	111	105,77 €
MBIT, LDA	1,27%	471	36,06 €
ONE VISION SERVIÇOS DE INFORMÁTICA E MULTIMÉDIA, LDA	1,23%	110	107,34 €
COPIALTA REPRESENTAÇÕES LDA	1,08%	203	50,05 €
UWE SCHARF	1,05%	43	171,90 €
NETZONE, LDA	1,02%	319	47,29 €
VGER - SOLUÇÕES TECNOLÓGICAS, LDA.	0,82%	26	177,08 €
MINHOMÁTICA-Form Prof e Inform Lda	0,75%	180	38,39 €
Manuel Fernando da Costa Campos	0,71%	124	33,22 €

Table 7: Ten customers with the highest share of customer.

When it comes to selling it is quite common that the more expensive the product the higher the revenue. As we do not have information about the revenue for product we will make this assumption.

In the month of January there were 1.006 different customers contributing to the company's results. But only 15% of these customers were responsible for around 56% of the turnover. Meaning that losing one of these clients would reflect as a 2.030€ loss in the turnover, in average. Or losing one of these customers is the same as losing 92 of the worse customers (in terms of share of customer).

VGER and MBIT most important products in terms of value

How does the different management policy really work? What does it mean? Where do we find that information? To answer these questions we will confront and compare two

specific customers: VGER – Soluções Tecnológicas, Lda and MBIT, Lda; from now on they will be referred simply as VGER and MBIT. These customers were chosen for their average product price, having VGER the highest value and MBIT the lowest from the group of customers with the share customer hit.

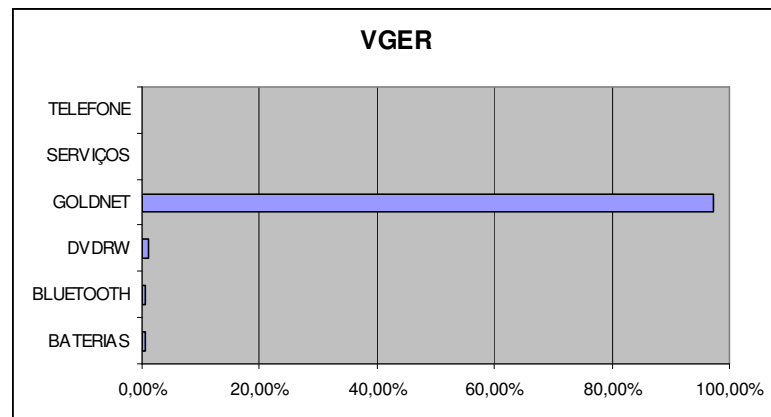


Chart 17: The relative importance of each product for the customer VGER in terms of value

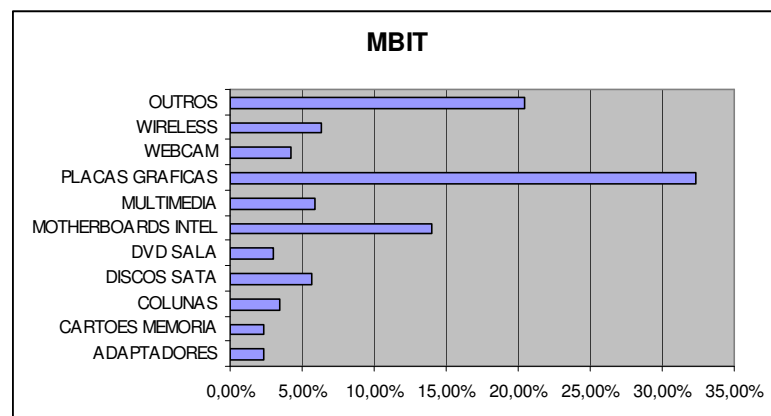


Chart 18: The relative importance of each product for the customer MBIT in terms of value

The product GOLDNET has an almost 100% relative importance in terms of value for the customer VGER, this means that summing up the value of all orders this product contributes with around 100% to the total. On the other side MBIT has a more widespread distribution of product relative importance in terms of value. VGER and MBIT are obviously very different customers in terms of product needs for their own

businesses; although they are both of extreme importance for Introduxi's turnover. If Introduxi wishes to maintain or increase the loyalty of these clients she should focus on their needs and not in what she wants to sell to them. It is important to invest in personalized mailing lists, personalized bonuses, to make clear to these customers that Introduxi desires to help them in improving their business and negotiating how it is possible.

VGER and MBIT most important products in terms of navigation performance

How does each customer behave navigating Introduxi's website? The only similitude between these two are: the average time spent per visit and the browser used. All the other measures are significantly different revealing different navigation patterns and consequently different needs.

Recency is defined as the time elapsed since the last visit by a visitor.

	VGER	MBIT
Pageviews	556	2299
No visits	24	150
Avg pageviews per visit	23,17	15,33
Frequency of visits (per day)	0,77	4,84
Recency of time	0	0
Avg time spent per visit	0:18:52	0:18:07
Avg time between visits (days)	1,22	0,19
Browser	Mozilla	Mozilla

Table 8: Confronting performance navigation results between a top share customer and a top buyer customer

What if we could know which specific products these clients are visualizing in the website and if they correspond to the ordered products?

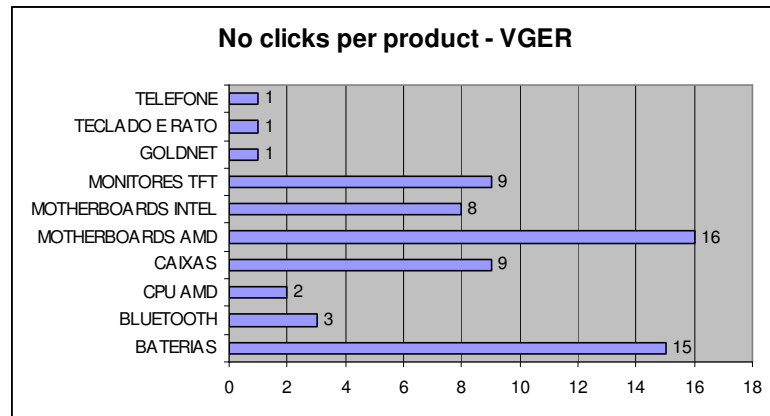


Chart 19: Number of clicks per product for the customer VGER

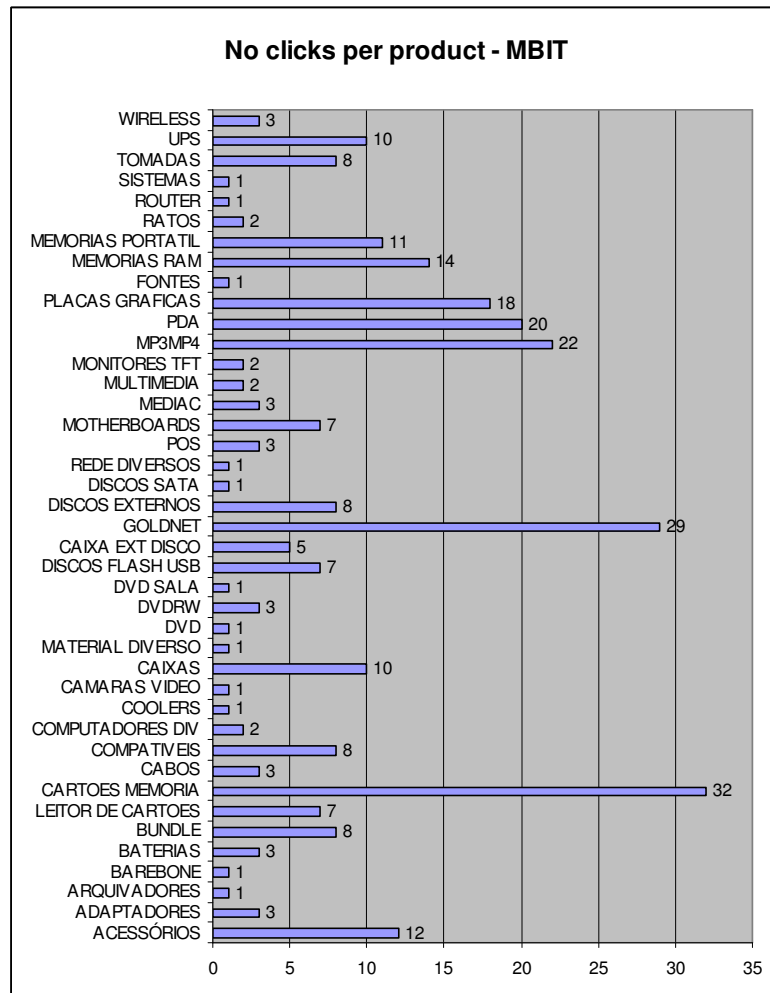


Chart 20: Number of clicks per product for the customer MBIT

In the case of VGER it is visible a preponderant interest in three products (Monitores TFT, Motherboards and Caixas) that are not ordered. All the other visualizations have been transformed into orders, with the exception of CPU AMD and Teclado e Rato. It is interesting to note that the most ordered product: GOLDNET, was visualised only once, although GOLDNET computers with different specifications have been ordered. To be more exact it was ordered 11 distinct specifications.

In the case of MBIT there is a wider variety of products than in the case of VGER. But the same comparison takes to the results that follow. The most seen products are not the ones most ordered. This is the case of GOLDNET, MP3/MP4 and PDA, with a number of orders of 0, 2, 0, respectively. It is pretty relevant to understand why MBIT never buys the products GOLDNET and PDA (both of them with high average price) when they show so much interest in visualizing them.

It was sold 26 Motherboards in the month of January, but there were only 7 clicks on the link that shows this product specification.

Summarizing...

We have first identified the most valuable customers, and then we have distinguished them by product money value and by their needs. Finally we have devoted ourselves to discover customer needs expressed not by their orders but their navigation options. With this action we have managed to interact with the client in a very discreet way, so discreet that he isn't even aware that this interaction happened. Now we may anticipate some of the consumers' needs and use this information not only for increase Introduxi's ROI but also to help clients to perform better.

The final step is to customize Introduxi's relationships with their customers, knowing which mailings should be sent to who, which customers should be more cherished, how can we save time and money to the client by making first available what he needs. This is the way for more ROI to the company.

5.2 Market Segmentation

The understanding of the customers' behaviour online helps to create a relationship with the customer. The desire to create and sustain long-term customer relationships that add value to the customer and the company has led to the development of customer relationship programs (**CRM**) (Turban et al., 2008, pp. 606). Not all the companies are ready to start a CRM program, it is necessary to implement a set of techniques in order to be able to make interactions with the client in a customized basis (Rayport , 2003).

“**Market segmentation** is the process of dividing a customer market into logical groups for conducting marketing research, advertising and sales.” (Turban et al., 2008, pp. 173) The way market segmentation is done is not the same way for the Business-to-Consumer (B2C) commerce as for the **Business-to-Business (B2B)** commerce (Turban et al., 2008 , pp. 179). The B2C refers to individual users while the B2B refers to corporate purchasers (Rayport et al., 2003, pp. 4, 5). Either way understanding the way the buying decision process is structured possibilities the selling company to influence the buyers' decision (Turban, 2008, pp. 157). The market segmentation provides a better response rate and the cost of reaching the segments is lower than the marketing directed to the masses (Turban, 2008, pp. 163).

If a company wishes to provide a strong total customer experience it needs to possess the adequate data on its customers (Seybold, 2002, pp. 114). **Database marketing** uses databases on customers to generate personalized relationships with those customers or customers to be (Hughes, 1994). Also associated with database marketing are statistical methods, data mining techniques and large repositories to keep the huge amount of necessary data (data warehousing) (Turban et al., 2008, pp.173). The process of extracting useful information from the databases in order to support better business decision making is denominated as **business intelligence** (Turban et al., 2008, pp. 178).

To proceed to a market segmentation analysis on the Introduxi's customers it was necessary to build a database. This database was built using click stream behaviour from the log files and joining some selling information from the shopping cart database.

The process of building the referred database was pretty complex and implied a considerable amount of hard-working.

The database reunites data both from log queries and selling data which was made available by the company in “xls” format.

Firstly the log was queried in order to obtain values for certain variables per customer (these variables will be explained later). Customers identified only by IP were not considered for the database. This decision was taken for two reasons. One of the reasons is that if a customer wants to buy he has to log in; it is important to remember that only customers who log in can visualize prices as well as particular links of the website. The other reason is that the selling data contains only data on the customers who actually bought something during the observation period.

Secondly the selling data was cleaned from the classic channel registries, leaving for analysis the net channel registries.

Finally log data and selling data had to be joined together by one common variable: the customer identification. All this steps led to the creation of a whole new database, which included the variables considered necessary and relevant for defining market segmentation.

The variables used in the database for market segmentation have their origin whether in the log data or in the selling data. The variables chosen from the log data were the following:

→ No pageviews – The number of pageviews corresponds to the total number of visualized pages in the observation period; this variable permits to identify how many pages were visualized per customer in the observation period.

→ Avg pageviews per session – The number of average pageviews per session is the ratio of the total number of visualized pages to the total number of sessions; this variable allows knowing how many pages are visualized, in average, every time a customer initializes a session.

→ Avg time per session – The average time per session is the ratio of total time spent on the site to the total number of sessions; this variable gives the information of how much time, in average, a customer spends on the website every time he initializes a session.

The variables chosen from the selling data were the following:

→ Customer share – The customer share corresponds to the ratio of customers' turnover to the total turnover; this variable permits to quantify how much each customer contributes to the company's turnover.

→ No units – The number of units is the total number of units ordered; this variable gives the information of how many units each customer ordered in the observation period.

→ Avg price per order – The average price per order corresponds to the ratio of total turnover to the total number of orders; this variable allows knowing the average price of the items ordered every time a customer makes an order.

The variables: No pageviews, Avg pageviews per session and Avg time per session reflect the customer performance on the site; usually a bigger interest on the site is related to a high probability of buying.

The clusters were determined using the SPAD software versão 5.5 (Système Pour Analyse de Données). The SPAD is a computer program used for statistical analysis. This program was chosen because it was available at the university. It was determined 7 clusters or seven segments, aggregating different customers.

Clusters:

→ **Cluster 1: “Non-interesting & bad performance”** (557 customers) – aggregates the non interesting customers (at least at the time of the analysis), the ones which visit the

site once in a while and buy very seldom. This cluster differs from the cluster number 5 for the worse performance in terms of navigation.

	CLUSTER CHARACTERISTICS - Average values per cluster					
	Customer share	No pageviews	No units	Avg pageviews per session	Avg time per session	Avg price per order
	(1)	(2)	(3)	(4)	(5)	(6)
Overall Average	0,099	469,438	31,791	21,189	952,000	44,960
Cluster 1 (557 customers)	0,045	317,346	18,106	16,373	659,000	29,570
Cluster 2 (64 customers)	0,206	337,406	13,391	17,775	829,000	162,163
Cluster 3 (8 customers)	0,210	131,250	1,875	22,759	868,000	678,303
Cluster 4 (10 customers)	0,035	878,600	11,600	76,701	4824,000	25,375
Cluster 5 (197 customers)	0,056	429,949	18,320	31,876	1580,000	32,958
Cluster 6 (140 customers)	0,214	1000,550	78,921	22,822	1001,000	34,921
Cluster 7 (30 customers)	0,620	1309,533	208,367	21,160	1029,000	43,959

Table 9: Average results per cluster considering the variables chosen for market segmentation

→ **Cluster 2: “Good share & Poor navigation performance”** (64 customers) – aggregates customers with the second best average price per order, after the customers from cluster 3. These customers have a very good customer share but a poor navigation performance, their results for the number of pageviews, the number of units ordered and the average pageviews per session are all below the overall average.

→ **Cluster 3: “Big value orders”** (8 customers) – aggregates the customers that distinguish from all the others for the value of their orders, these customers present the highest average price per order.

→ **Cluster 4: “Online all day long”** (10 customers) – aggregates the customers with the best results in terms of average time per session but not as good in terms of visualized

pages. These customers seem to spend most of their day online, clicking on the links of the website.

→ **Cluster 5: “Non-interesting”** (197 customers) – aggregates the customers with a very low customer share even though they present reasonable navigating performance, they spent a lot of time online and visualize a considerable amount of pages per session, but they do not convert this interest into orders. This cluster has a slightly better average price per order than the cluster number 1.

→ **Cluster 6: “Good share & Good performance”** (140 customers) – aggregates customers with a very good customer share. Their results for the number of visualized pages are considerably good as well as the number of units ordered. This cluster distinguishes from the cluster number 2, mainly for the higher number of visualized pages and number of units ordered.

→ **Cluster 7: “Top share customers”** (30 customers) – aggregates top customers in terms of share and in terms of visualized pages. They are characterized for the highest number of orders and also highest number of visualized pages.

The Chart 21 pretends to evidence the disparity of results of the variables chosen to conduct market segmentation. On the x axis are represented the 6 variables: (1) Customer share, (2) Number of pageviews, (3) Number of units, (4) Average pageviews per session, (5) Average time per session and (6) Average price per order. On the y axis are the numeric results for each variable per cluster. Each cluster is distinguished by a different colour as can be seen in the legend.

The purpose of segmenting customers according to their web behaviour and web purchases is to make the web site’s marketing actions more profitable and obtain competitive advantages. (Ruiz Mafé et al., 2006)

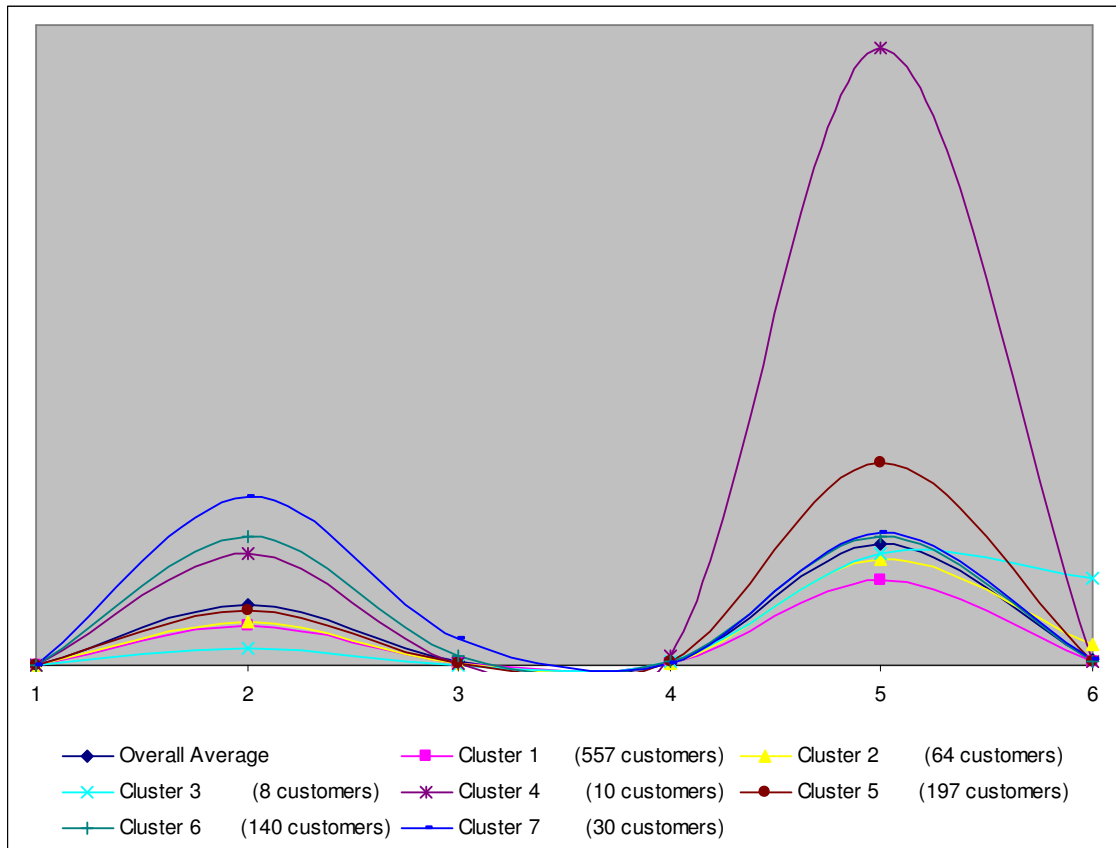


Chart 21: Average results per cluster considering the variables chosen for market segmentation

The analysis made in chapter 4.2.3 Log analysis and 4.2.4 Log analysis - Customer Session was critical for measuring the effectiveness of Introduxi's web site and determining improvement actions needed. But what if we could measure the effectiveness of an e-commerce site by different segment of customers? In this chapter we will compute Relative Contact Efficiency and Conversion Efficiency measures, this time by each segment of customer determined according to their web performance and selling registries. It was necessary to rewrite some queries. Again it was determined the number of customer sessions but by cluster.

The measure Relative Contact Efficiency was determined as follows:

(Query no 5)

```

SELECT session.id
FROM session, user, usage, page, date
WHERE session.dimension_key = usage.session AND page.dimension_key = usage.page
AND date.dimension_key = usage.date AND user.dimension_key = usage.user
AND date.date_id BETWEEN '2008-01-01' and '2008-01-31'
AND user.identification in (...cluster...)
AND page.page_id like '%PORT%'
AND page.page_id like '%info%'

INTERSECT

SELECT session.id
FROM session, user, usage, page, date
WHERE session.dimension_key = usage.session AND page.dimension_key = usage.page
AND date.dimension_key = usage.date AND user.dimension_key = usage.user
AND date.date_id BETWEEN '2008-01-01' and '2008-01-31'
AND user.identification in (...cluster...)
AND page.page_id like '%encomendar%'

```

Note: For further detail on this query please check on appendix 1.

$$Relative_Contact_Efficiency_{\substack{Cluster_y \\ product_x}} = \frac{Query5_{\substack{Cluster_y \\ product_x}}}{No_customer_sessions_{Cluster_y}}$$

Relative Contact Efficiency		Cluster						
		1	2	3	4	5	6	7
Type of product	ACESSÓRIOS	25,10%	16,54%	28,57%	58,33%	36,36%	24,96%	18,64%
	COMPONENTES	38,23%	44,88%	14,29%	25,00%	76,86%	56,45%	52,54%
	COMPUTADOR	4,58%	19,69%	100,00%	0,00%	9,37%	9,38%	7,63%
	CONSUMÍVEIS	3,96%	6,30%	0,00%	33,33%	6,61%	4,36%	3,39%
	IMAGEM	3,33%	5,51%	14,29%	16,67%	5,23%	4,36%	3,81%
	PERIFÉRICOS	34,27%	40,94%	42,86%	100,00%	52,62%	39,53%	33,05%
	REDES/COMUNICAÇÃO	13,13%	10,24%	0,00%	25,00%	17,91%	15,41%	12,71%
	SOFTWARE	1,25%	9,45%	0,00%	0,00%	2,20%	4,02%	2,54%

Table 10: Relative Contact Efficiency results per cluster and by type of product

Analysing Table 10 and confronting it with the results from Table 9, we can conclude that:

→ **Cluster 1: “Non-interesting & bad performance”** (557 customers) – the action pages or the type of products that are relatively more important in the site for these customers are the ones with the lowest average prices

→ **Cluster 2: “Good share & Poor navigation performance”** (64 customers) – the type of products Componentes and Periféricos are the most important within the site, although the majority of this type of products are cheap, this cluster leans to the products of greater value within this range

→ **Cluster 3: “Big value orders”** (8 customers) – the most important type of product for this cluster are the Computador, this makes a lot of sense since they are known for the incredibly high average price of their orders

→ **Cluster 4: “Online all day long”** (10 customers) – the type of products relatively more important to the customers within this cluster are Periféricos, they visualize more pageviews per session than any other individuals, having an interesting relative contact efficiency result of 100%

→ **Cluster 5: “Non-interesting”** (197 customers) – as for cluster 1 the type of products that are relatively more important in the site for these customers are the ones with the lowest average prices

→ **Cluster 6: “Good share & Good performance”** (140 customers) – the type of products Componentes and Periféricos are the most important within the site for this cluster; it offsets the low value of the products ordered with the volume of quantities ordered

→ **Cluster 7: “Top share customers”** (30 customers) – the action pages relatively more important in the site are the ones concerning the type of product Componentes, again a type of product with a low average price; this cluster totally wins the ordering championship, achieving via quantity the utmost customer share

The measure Conversion Efficiency was determined as follows:

(Query no 6)

```
SELECT session.id  
FROM session, user, usage, page, date  
WHERE session.dimension_key = usage.session AND page.dimension_key = usage.page
```

```

AND date.dimension_key = usage.date AND user.dimension_key = usage.user
AND date.date_id BETWEEN '2008-01-01' and '2008-01-31'
AND user.identification in (...cluster...)
AND page.page_id like '%PORT%'
AND page.page_id like '%info%'

INTERSECT

SELECT session.id

FROM session, user, usage, page, date

WHERE session.dimension_key = usage.session AND page.dimension_key = usage.page
AND date.dimension_key = usage.date AND user.dimension_key = usage.user
AND date.date_id BETWEEN '2008-01-01' and '2008-01-31'
AND user.identification in (...cluster...)
AND page.page_id like '%PORT%'
AND page.page_id like '%comprar%'

INTERSECT

SELECT session.id

FROM session, user, usage, page, date

WHERE session.dimension_key = usage.session AND page.dimension_key = usage.page
AND date.dimension_key = usage.date AND user.dimension_key = usage.user
AND date.date_id BETWEEN '2008-01-01' and '2008-01-31'
AND user.identification in (...cluster...)
AND page.page_id like '%encomendar%'

```

Note: For further detail on this query please check on appendix 1.

$$Conversion_Efficiency_{product_x}^{Cluster_y} = \frac{Query6_{product_x}^{Cluster_y}}{Query5_{product_x}^{Cluster_y}}$$

Conversion Efficiency		Cluster						
		1	2	3	4	5	6	7
Type of product	ACESSÓRIOS	56,43%	66,67%	50,00%	85,71%	50,00%	50,34%	65,91%
	COMPONENTES	65,40%	42,11%	100,00%	0,00%	60,22%	49,85%	62,10%
	COMPUTADOR	27,27%	60,00%	71,43%	0,00%	17,65%	19,64%	44,44%
	CONSUMÍVEIS	57,89%	75,00%	0,00%	50,00%	79,17%	73,08%	50,00%
	IMAGEM	62,50%	71,43%	100,00%	0,00%	57,89%	42,31%	55,56%
	PERIFÉRICOS	62,31%	51,92%	0,00%	58,33%	61,78%	48,31%	61,54%
	REDES/COMUNICAÇÃO	59,52%	38,46%	0,00%	33,33%	47,69%	40,22%	63,33%
	SOFTWARE	91,67%	83,33%	0,00%	0,00%	37,50%	41,67%	33,33%

Table 11: Conversion Efficiency results per cluster and by type of product

Table 11 results are comparable with the customer sessions results, but for which cluster we have determined the number of customer sessions. Analysing Table 11 and confronting it with the results from Table 9, we can conclude that:

→ The overall results for the conversion efficiency measure are relatively good. A wide number of types of products have results above 50% and most clusters have average results above 40%.

→ The type of product Componentes and Imagem have a conversion efficiency result of 100% among the customers of the cluster 3. Meaning that every click in an action page of these two products performed by cluster 3 customers delivers in a conclusion of the buying process.

→ There are some results that are not surprising considering the previous analysis; this applies to the following clusters: 2 “*Good share & Poor navigation performance*”, 3 “*Big value orders*”, 6 “*Good share & Good performance*” and 7 “*Top share customers*”. When studying the impact of each page in the success of the site I wouldn’t expect less than this conversion efficiency results from the customers of these clusters.

→ However clusters 1 “*Non-interesting & bad performance*” and 5 “*Non-interesting*” showed an unexpectedly good conversion efficiency results for most type of products. It could be interesting to develop a targeted marketing campaign for the customers of these clusters. The purpose would be to increase these customers interest in higher added value products similarly to cluster 2 or to raise the number of orders among these customers similarly to cluster 6. Since the behaviour of these two clusters is very similar it would be better to test this marketing campaign in only one of the clusters. It could make more sense start with cluster 5 “*Non-interesting*” since its navigation performance results are better than the ones from cluster 1 “*Non-interesting & bad performance*”.

5.3 Combining sales and success metrics

It is time to make a synthesis by aggregating all the numbers obtained till this point. In this chapter we will combine sales numbers obtained from the selling registries or accounting results with the success metrics obtained from the log. The sales numbers consider three determinant variables for the success of the business like quantities sold, average price of the products and turnover of each product. The success metrics come from the calculation of the contact efficiency, relative contact efficiency and conversion efficiency measures considering the entire log.

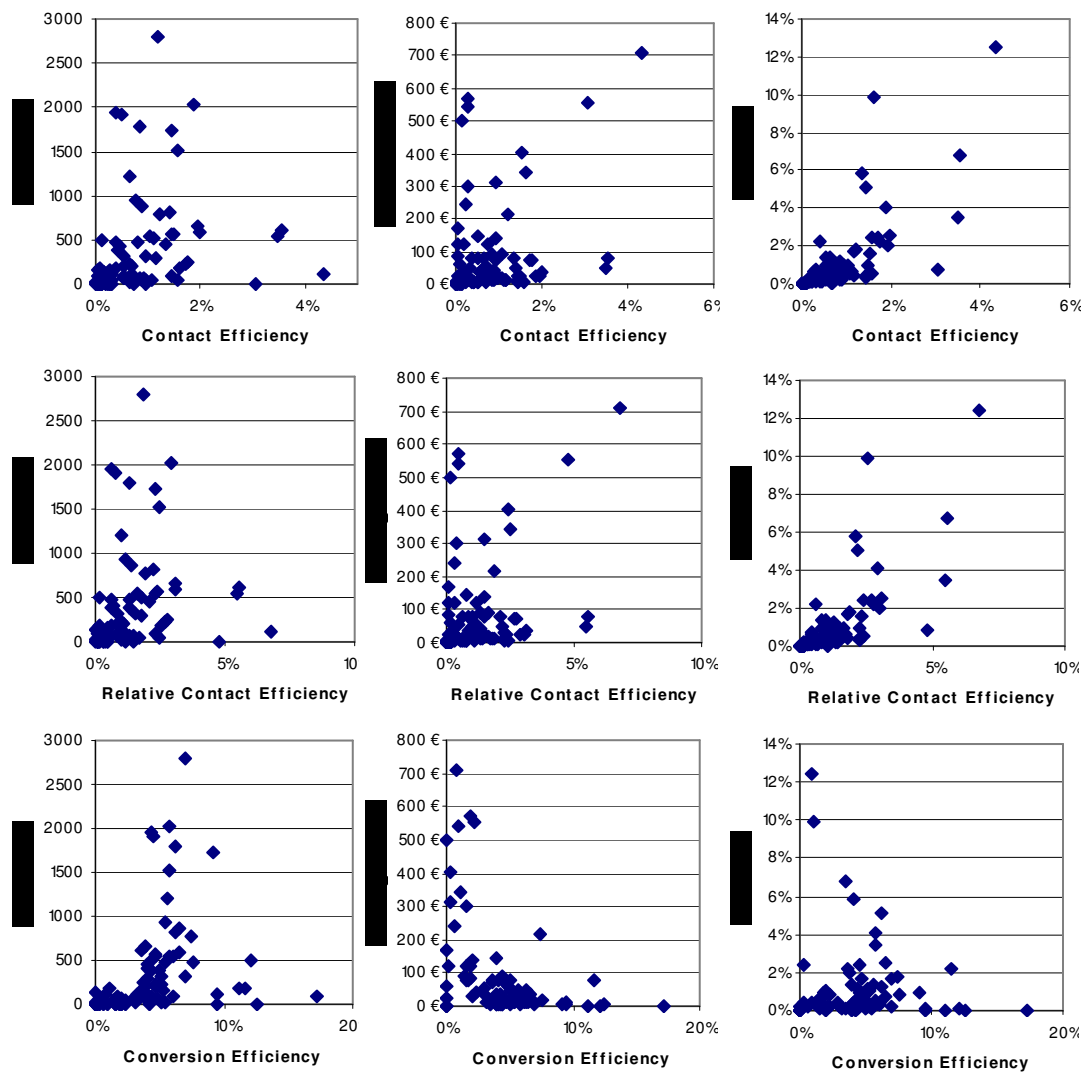


Chart 22: Combining sales and success metrics per family of products

Chart 22 intends to highlight the families of products that behave as an outlier when confronting the sales variables and the success metrics results, each point represents a family of products. There are 81 families of products. On the x axis are represented the success metrics: contact efficiency, relative contact efficiency, conversion efficiency. On the y axis are the sales variables: quantities, average price and turnover.

As we can see in Chart 22 there are some products behaving as outliers that we should draw our attention to, let us look at some examples:

- The Plasma/LCDTV present good contact efficiency results but a significantly low result for the turnover (a result below the average which is around 1,2%)
- The laptops have got the highest relative contact efficiency results as well as the highest average price
- The DVD burners stand with very good conversion efficiency results as well as with a considerable high price

5.4 Deployment Plan

The case study in this dissertation was built on several different steps. First we had to find a company available for delivering their data for analysis and then we had to collect the data. After collection, the data was stored in a data warehouse and pre-processed. Data was first analyzed as a whole and we decided how to model it. We used queries for obtaining the numeric results. Finally we analysed these results trying to find the cases that could bring some improvement to the web site.

We should take a further look to Chart 22 and outline the situations that deviate from the normal (here the normal are the cases coincident with the findings of the dissertation):

→ It seems odd to find a product with high contact efficiency results when having low turnover. We have seen that the most popular products are the ones with the highest turnover. By the turnover rate of a certain product we mean the ratio of a certain product sales to the total amount of sales (including all products traded).

→ In the case of the relative contact efficiency measure it is unusual to have high results for products with high average prices.

→ It is also a little awkward to see high conversion efficiency for a product with high price. We have seen that the products that sell the most are the ones with the lowest average prices.

What takes to these exception situations? Are these exceptions positive or negative? Can we improve the success of the web site?

It would be good that the products with high contact efficiency would have high turnover, this means better net results to the company. To conclude this we are assuming that the products with higher prices imply bigger revenues. A product with low turnover means that it doesn't sell much even though it is successful. What can we do to make it sell?

We have also seen that the products that are more successful are the ones with relatively low prices. We should also make our best to allow high average products to be successful also.

Finally we have observed that the products that sell are the ones with the lowest average prices. It is very important to also sell the products that have higher average prices because these products imply better net results to the company. We are again assuming that the products with higher prices imply bigger revenues.

6 Conclusions

The final considerations on this dissertation will follow the same order as the dissertation itself.

The case study approached in this dissertation was constructed on several different steps. First we had to find a company available for delivering their data for analysis and then we had to collect the data in different moments of time. It was necessary to meet with the person in charge of the web site to understand the particularities of the business and the company's web site. After the collection the data was stored in a data warehouse and pre-processed. Data was first analyzed as a whole and we decided how to model it. We used queries for obtaining the numeric results. Finally we looked at these results and detected/highlighted the cases that could use some improvement according to the purpose of the project.

6.1 *Preliminary Analysis*

The first analysis made to the selling registries revealed the products with the highest weight in quantities are systematically the ones with the lowest average prices and the products with the highest weight in turnover are the ones with the utmost average price. The turnover is also dependent on a small percentage of products which might not be a good option. Also to notice that the turnover reflects the quantity $\times$ price, it would have been better to confront Net results, but we did not have access to this information.

The analysis to the log file showed that the number of page requests is much higher than the number of sessions, indicating that the users of the website click a lot each at session. The behaviour over the week of the number of page requests and orders very much follows a 9 to 5 office schedule. The traffic hours in terms of page requests are from 10 a.m. to 12 p.m. and from 3 to 6 p.m.; in terms of orders the most popular time of the day is from 10 a.m. to 12 p.m. and at 6 p.m. The numbers show that the users who click the most in the web site are not the ones who buy the most.

The analysis to the log was made considering 980.945 page requests, 74.465 sessions, 47.590 active sessions, 2.302 customer sessions and 45.288 non-customer sessions.

6.2 Contact Efficiency

The overall results for the contact efficiency measure are low. This does not mean that the site is not successful; it has to be taken in consideration that there are 81 action pages. In other words, a user is not likely to visit such a number of different products every session.

The relative contact efficiency is a measure appropriate for sites with many action pages, which is the case, or a large number of inactive sessions, which is also the case. The results of this measure are not so good, what again is understandable if we think that this measure expresses the relative importance of each action page within the site. When we have 81 action pages we can not expect that every page is a hit.

6.3 Conversion Efficiency

The results for the conversion efficiency measure on the overall are worse than those obtained for the Relative Contact Efficiency measure, with the exception of Consumiveis and Software. Of course some action pages are more successful in guiding the user towards the site's goal than others.

The first analysis made to the log was followed by the same analysis but for the customer sessions. A customer session is a session in which the user has achieved the site's goal.

It was showed that the results of the measures are far better using the customer session log. But the most popular types of products for the Relative Contact efficiency results in the customer log are still the most popular when considering the customer session log.

When comparing the results for the measure Conversion efficiency (entire log vs customer session log) it is seen that when considering the customer session log the results are on average around 50%. This means that, on average, each action page has a 50% success on leading the customer to the target page. This is a pretty good result for the conversion efficiency measure within the buying customers.

If we order (by importance) the conversion efficiency results of both analysis (entire log vs customer session log) it is interesting to notice that almost all the products changed their order of importance, with the exception of the type of product Computador that maintained the last position in both the analysis. The most important products in terms of conversion efficiency results, despite their order, are still the same four (Acessorios, Componentes, Consumiveis and Perifericos) in the customer session log analysis.

When working on the buying process we defined a certain number that needed to be accomplished to achieve the site's goal. We considered that one of these steps should be an action page (step 2) and other the target page (step 5). It is important to notice that it is possible to jump over step 2, when adding a specific product to the shopping cart. After selecting the kind of product the user is interested in (step 1), instead of having a closer look to the specific product he has chosen (step 2) he immediately adds it to the shopping cart (step 3).

6.4 Focusing on top costumers

We carried out a business intelligence analysis over two specific customers in order to show, in practical terms, what kind of information can be extracted from databases that might support better decision making.

To choose the customers in whom to base this analysis we confronted top results in terms of customer share, quantities ordered and average prices of those orders. Then we studied the relative importance of each product for those two customers as well as some navigation performance measures obtained from the log; measures like number of pages visualized, number of sessions, frequency of visits, average time spent per visit and

average time between visits. With this action we have managed to interact with the client in a very discreet way, so discreet that he isn't even aware that this interaction has happened. Now we may anticipate some of the consumers' needs and use this information not only for increasing Introduxi's ROI but also to help clients perform better.

We noticed that a great interest in a product does not imply buying it. We have also noticed that customers with very good customer share behave very differently in terms of navigation patterns as well as buying orders.

If Introduxi wishes to maintain or increase the loyalty of these clients she should focus on their needs and not in what she wants to sell to them. It is important to invest in personalized mailing lists, personalized bonuses, to make clear to these customers that Introduxi desires to help them in improving their business and negotiating how it is possible.

The final step is to customize Introduxi's relationships with their customers, knowing which mailings should be sent to whom, which customers should be more cherished, how can we save time and money to the client by making first available what he needs. This is the way for more ROI to the company.

6.5 Market Segmentation

After completing the business intelligence analysis and understanding the difference of behaviour in terms of orders and navigation patterns we understood that we needed to compute the success measure again, this time to small groups of customers that would be homogeneous inside that group.

To proceed to a market segmentation analysis on the Introduxi's customers it was necessary to build a database. This database was built using click stream behaviour from the log files and joining some selling information from the shopping cart database.

The variables used in the database for market segmentation have their origin whether in the log data or in the selling data. The variables chosen from the log data were the following: the number of pageviews, the number of average pageviews per session, the average time per session, the customer share, the total number of units ordered, the average price per order.

We have found 7 clusters, namely: Cluster 1: “Non-interesting & bad performance” (557 customers), Cluster 2: “Good share & Poor navigation performance” (64 customers), Cluster 3: “Big value orders” (8 customers), Cluster 4: “Online all day long” (10 customers), Cluster 5: “Non-interesting” (197 customers), Cluster 6: “Good share & Good performance” (140 customers), Cluster 7: “Top share customers” (30 customers).

6.6 Site’s success per segment

We computed Relative Contact Efficiency and Conversion Efficiency measures, this time by each segment of customer determined according to their web performance and selling registries. Again it was determined the number of customer sessions but by cluster. The interpretation of the measures per cluster confronted with average results of the database created, allowed us to better understand navigation pattern behaviours and consequently buying options. From now on we can define marketing actions suitable to each cluster profile, instead of overloading with non-interesting advertising. This kind of approach should turn out more profitable and bring competitive advantages over a non-personalized addressing.

Following all these calculations we made a synthesis chart which aggregates all the numbers obtained till this point. We combined sales numbers obtained from the accounting results with the success metrics obtained from the log. The sales numbers considered three determinant variables for the success of the business like quantities sold, average price of the products and turnover of each product. The success metrics numbers came from the calculation of the contact efficiency, relative contact efficiency and conversion efficiency measures considering the entire log.

6.7 Sales versus Efficiency

The synthesis chart intends to highlight the families of products that behave as outliers when confronting the sales variables and the success metrics results. Analysing the synthesis chart we detected the following peculiar situations: products with high contact efficiency results when having low turnover, products with high relative contact efficiency and high average prices, products with high conversion efficiency and high prices.

6.8 Main contributions

In this dissertation we have applied the success measures to a real case, with real data. We have observed that products with relatively high contact efficiency can have low conversion efficiency and vice versa. This highlights the pages which need more attention and the ones which are already successful.

We have determined the success measures within the customer sessions. We have observed that within these sessions we have very high conversion efficiency results, meaning that almost all pages are successful.

We have segmented the clients according to sales and navigation behaviour and studied the success of the site by sector. This enables the observation of the success of parts of the site for different user groups. We can, for example, observe that the action pages for the type of product Computador are more successful than they seemed if we consider the cluster who buys more. We can therefore design different action pages to different customer segment.

Also, in this dissertation, we have compared the sales data with the success data. We conclude that, despite we observe major tendencies; some products' pages escape the general laws. This is the case of the laptops because this product is the most successful in terms of relative contact efficiency results and has got the highest average price. This

is a little odd since we have concluded that in the case of the relative contact efficiency measure it is unusual to have high results for products with high average prices. This is also the case of the DVD burners which stand with very good conversion efficiency results as well as with a considerable high price. It is also a little awkward to see high conversion efficiency for a product with high price. We have seen that the products that sell the most are the ones with the lowest average prices.

6.9 Future work

First we would like to collect contact and conversion efficiency measures along a period of time and determine high and low values that can be used to monitor the success of the site. At the same time we are interested in performing a similar study for each of the identified segments.

It would be interesting to relate sudden variations in the success of the site and of its pages to decisions made by the marketing team, or to outside events. For example, if a new campaign increases contact but reduces conversion, this could be studied.

We want to define a method for detecting interesting outliers in the relation between sales data and success data. If a product becomes an outlier maybe it should deserve a closer attention.

Finally we would like to implement a tool that collects data and continuously compiles the measures results and makes it available to the site's management. This way management could make decisions on real time based on how the customers are behaving on their web site.

Bibliography

Papers

Liao, C., P. Palvia and H. Lin (2006), "The roles of habit and web site quality in e-commerce", *International Journal of Information Management* 26, pp.469-483.

Norguet, J., E. Zimányi and R. Steinberger (2006), "Improving Web sites with web usage mining, web content mining and semantic analysis", *Lecture Notes in Computer Science* 3831, pp. 430-439.

Spiliopoulou, M. and C. Pohle (2001), "Data mining for measuring and improving the success of web sites", *Data Mining and Knowledge Discovery*, 5, pp.85-114.

Berendt, B. (2002a), "Using site semantics to analyze, visualize, and support navigation", *Data Mining and Knowledge Discovery*, 6, pp.37-59.

Berendt, B. (2002b), "Detail and Context in Web Usage Mining: Coarsening and visualizing sequences", *Lecture Notes in Artificial Intelligence* 2356, pp. 1-24.

Anand, S. S., M. Mulvena and K. Chevalier (2004), "On the deployment of web usage mining", *Lecture Notes in Artificial Intelligence* 3209, pp. 23-42.

Anand, S. S. and B. Mobasher (2005), "Intelligent techniques for web personalization", *Lecture Notes in Artificial Intelligence* 3169, pp. 1-36.

Liang, T. and H. Lai (2002), "Effect of store design on consumer purchases: Van empirical study on-line bookstores", *Information & Management* 39, pp. 431-444.

Reichheld, F. F. and P. Schefter (2000), "E-loyalty: your secret weapon on the web", *Harvard Business Review*, July-August 2000, pp. 105-113.

Büchner, A. G. and M. D. Mulvenna (1998), "Discovering internet marketing intelligence through online analytical web usage mining", *SIGMOD Record*, pp. 54-61.

Facca, M. L. and P. L. Lanzi (2005), "Mining interesting knowledge from weblogs: a survey", *Data knowledge engineering* 53, pp. 225-241.

Mobasher, B., R. Cooley and J. Srivastava (2000), "Automatic personalization based on web usage mining", *Communications of the ACM*, Vol. 43, No 8, pp. 142-151.

Srivastava, J., R. Cooley, M. Deshpande and P. Tan (2000), "Web usage mining: discovery and applications of usage patterns from web data", *ACM SIGKDD*, Vol. 1, Issue 2, pp. 12-23.

Berthon, P., L. F. Pitt and R. T. Watson (1996), "The world wide web as an advertising medium", *Journal of advertising research*, January/February 1996 pp. 43-54.

Cooley, R., B. Mobasher and J. Srivastava (1999), “Data preparation for mining world wide web browsing patterns”, Journal of knowlwdge and information systems, pp. 5-32.

Ho, J. (1997), “Evaluating the World Wide Web: a global study of commercial web sites”, Journal of computer mediated communication.

Shearer, C. (2000), “The CRISP-DM Model: The New Blueprint for Data Mining”, Journal of Data Warehousing, Vol. 5, No 4, pp. 13-22.

Fayyad , U., G. Piatetsky-Shapiro and P. Smyth (1996a), “The KDD Process for Extracting Useful Knowledge from Volumes of Data”, Communications of the ACM, Vol. 39, No 11, pp. 27-34.

Fayyad, U., G. Piatetsky-Shapiro and P. Smyth (1996b), “From Data Mining to Knowledge Discovery in Databases”, American Association for Artificial Intelligence, Fall 1996, pp. 37-54.

Kim, J. K., Y. H. Cho, W. J. Kim, J. R. Kim and J. H. Suh (2002), “A personalized recommendation procedure for Internet shopping support”, Electronic Commerce Research and Applications 1 (2002), pp. 301–313.

Zaiane, O. R., M. Xin, J. Han (1998), “Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs”, Advances in Digital Libraries.

Ruiz Mafé, C. and C. Lassala Navarré (2006), “Segmenting consumers by e-shopping behaviour and online purchase intention”, Journal of Internet Business.

Books

Hortinha, J. (2001), E-marketing - Um guia para a Nova Economia, Edições Sílabo.

Turban, E., D. King, J. McKay, P. Marshall, J. Lee and D. Viehland (2008), Electronic Commerce 2008 - A Managerial Perspective, Pearson Prentice Hall

Sterne, J. (2002), Web Metrics - Proven Methods for Measuring Web Site Success, Wiley Publishing, Inc.

Ledford, J. L. and M. E. Tyler (2007), Google Analytics 2.0, Wiley Publishing, Inc.

Travis, D. (2003), e-commerce usability – tools and techniques to perfect the on-line experience, Taylor & Francis

Reis, J. L. (2000), O Marketing Personalizado e as Tecnologias da Informação, Edições Centro Atlântico

Seybold, P. B., R. T. Marshak and J. M. Jeffrey (2002), The Customer Revolution – How to thrive when your customers are in control, Business Books

Windham, L. and K. Orton (2000), The Soul of the New Consumer – The attitudes, Behaviors, and preferences of e-customers, New York: Allworth Press

Bruner, R. E., L. Harden and B. Heyman (2001), NetResults.2 - Best Practices for Web Marketing, New Riders Publishing

Chaffey, D. (2004), E-Business and E-Commerce Management, Second Edition, Pearson Education

Rayport, J. F. and B. J. Jaworski (2003), introduction to e-commerce, McGraw-Hill

Siegel, J. and J. Shim (2003), Database Management Systems, Thomson

Sheth, J. N. and A. Parvatiyar (2000), Handbook of Relationship Marketing, Sage Publications, Inc.

Moroco, J. (2003), Análise Estatística - Com utilização do SPSS, Edições Sílabo

Hughes, A. M. (1994), Strategic Database Marketing – The Masterplan for Starting and Managing a Profitable, Customer-Based Marketing Program, Second Edition, McGraw-Hill

Reis, E. (1997), Estatística Multivariada Aplicada, Lisboa: Edições Sílabo

Kotler, P. (1997), Marketing Management – Analysis, Planning, Implementation and Control, NJ: Prentice-Hall, Englewood Cliffs

Software

Système SPAD pour Windows Version 5.5 (no date). Montreuil: CISIA-CRESTA

Appendix 1

Code	What does it mean?
structure	dw4introduxi.fact_table_structure
usage	dw4introduxi.fact_table_usage
content	dw4introduxi.fact_table_content
session	dw4introduxi.dimension_table_session
user	dw4introduxi.dimension_table_user
time	dw4introduxi.dimension_table_time
date	dw4introduxi.dimension_table_date
page	dw4introduxi.dimension_table_page
usage.date	dw4introduxi.fact_table_usage.dimension_table_date
usage.page	dw4introduxi.fact_table_usage.dimension_table_page
usage.time	dw4introduxi.fact_table_usage.dimension_table_time
usage.session	dw4introduxi.fact_table_usage.dimension_table_session
usage.user	dw4introduxi.fact_table_usage.dimension_table_user

Appendix 2

	Família	Family of Product	Product Code Reference
ACCESSORIES	CAIXAS E BOLSAS	CASES AND BAGS	ACE
	ADAPTADORES	ADAPTERS	ADP
	ARQUIVADORES	CD CASES	ARQ
	BATERIAS	BATTERIES	BAT
	CABOS	CABLES	CB
	CAIXA EXT DISCO	EXT. ENCLOSURE	DISCO
	DATA SWITCH	DATA SWITCH	DATASWITCH
	FICHAS	PLUGS	FICHAS
	MALA PORTATIL	LAPTOP CASES	MALAP
	MATERIAL DIVERSO	ACESSORIES	DIV
	MICROFONES	MICROFONES	MICROS
	MP3/MP4	MP3/MP4	MP3
	TAPETES	MOUSE PADS	TAP
	TOMADAS	OUTLET	TOMADA

	Família	Family of Product	Product Code Reference
COMPONENTS	BAREBONE	BAREBONE	BARE
	CAIXAS	CASES	CX
	CONTROLADORAS	CONTROLLERS	CTRL
	COOLERS	COOLERS	COOLER
	CPU	CPU	CPU
	DISCOS	HARD DRIVES	HD
	DRIVES DISQUETE	FLOPPY DRIVE	DRIVEDSK
	DRIVES DIVERSAS	DIVERS DRIVES	DRIVEDVS
	LEITOR DVD E HD	DVD BURNERS	DVDR
	FONTES	POWER SUPPLIES	PW
	LEITOR DE CARTOES	CARD READERS	CARDREADER
	MEMORIAS	MEMORIES	RAM
	MOTHERBOARDS	MOTHERBOARDS	MBA
	MOTHERBOARDS	MOTHERBOARDS	MBI
	PLACAS GRAFICAS	VIDEO CARDS	PLGRA
	PLACAS REDE	NETWORK CARDS	PLR
	PLACAS SOM	SOUND CARDS	PLSOM
	PLACAS VIDEO	TV CARDS AND VIDEO DEVICES	PLVD

	Família	Family of Product	Product Code Reference
DESKTOPS / LAPTOPS	COMPUTADORES	COMPUTERS	COMPDIV
	COMPUTADORES	COMPUTERS	COMPHP
	GOLDNET	GOLDNET	GOLD
	GOLDNET	GOLDNET	PC
	PORTATIL	LAPTOP	PORT
	SERVIDORES	SERVERS	SERVIDORES

	Família	Family of Product	Product Code Reference
CONSUMABLES	CANETAS CDRPEN	CD PEN	CANETACDRPEN
	CASSETES	CASSETS	CASSETTE
	CD	CD	CD
	COMPATIVEIS	COMPATIBLES	COMPATIVEL
	DISQUETES	FLOPPY DISCS	DISQUETES
	DVD	DVD	DVD
	PAPEL	PAPER	PAPEL
	TINTEIROS	INK CARTRIDGES	TINT
	TONNER	TONERS	TONNER

	Família	Family of Product	Product Code Reference
VIDEO	CAMARAS DIGITAIS	DIGITAL CAMERAS	CAMDIG
	CAMARAS VIDEO	CAMCODERS	CV
	CARTOES MEMORIA	MEMORY CARDS	CARTMEM
	DVD SALA	DVD PLAYERS	DVDSALA
	DVDPORT	PORTABLE DVD PLAYER	DVDPORT
	DVDRW SALA	DVD RECORDERS	DVDRWSALA
	PROJECTORES	PROJECTOR	PROJ

	Família	Family of Product	Product Code Reference
PERIPHERALS	BUNDLES	BUNDLES	BUNDLE
	COLUNAS	SPEAKERS	COL
	DISCOS EXTERNOS	EXTERNAL HARD DRIVES	HDEXT
	DISCOS FLASH USB	USB FLASH DRIVE	FLASHU
	IMPRESSORAS	PRINTERS	IMP
	JOYSTICKS	JOYSTICKS	JOYSTICKS
	MULTIMEDIA	MULTIMEDIA	MM
	PLASMA/LCDTV	PLASMA/LCDTV	MONTFT
	POS	POS	POS
	TECLADO	KEYBOARD	TEC
	RATOS	MOUSE	RATO
	UPS	UPS	UPS
	WEBCAM	WEBCAM	WEBCAMC

	Família	Family of Product	Product Code Reference
NETWORKING	BLUETOOTH	BLUETOOTH	BLTH
	FAX	FAX	FAX
	GPS	GPS	GPS
	HUB\SWITCH	HUB\SWITCH	HUBS
	MODEMS ANALOGICOS	ANALOGIC MODEMS	MODEMAN
	PDA	PDA	PDA
	REDE DIVERSOS	NETWORKING	HUBDIV
	ROUTER	ROUTER	ROUTER
	TELEFONE	PHONE	TELEF
	WIRELESS	WIRELESS	WIRELESS

	Família	Family of Product	Product Code Reference
SOFTWARE	ANTIVIRUS	ANTIVIRUS	ANTV
	SOFTWARE	SOFTWARE	SWS
	SOFTWARE	SOFTWARE	SWO
	SOFTWARE	SOFTWARE	SWD

Appendix 3

Queries

Number of page requests

```
select count (dw4introduxi.dimension_table_page.page_id)
from dw4introduxi.dimension_table_page, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-31'
```

Number of sessions

```
select dw4introduxi.dimension_table_date.date_id, count(distinct
dw4introduxi.dimension_table_session.id)
from dw4introduxi.dimension_table_date, dw4introduxi.dimension_table_session,
dw4introduxi.fact_table_usage
where dw4introduxi.dimension_table_session.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_session
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-31'

Group by dw4introduxi.dimension_table_date.date_id
order by dw4introduxi.dimension_table_date.date_id asc
```

Number of active sessions

```
select count (distinct dw4introduxi.dimension_table_session.id)
from dw4introduxi.dimension_table_session, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_page, dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_session.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_session
and dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
and page_id like '%info%'
```

Number of customer sessions

```
select distinct dw4introduxi.dimension_table_session.id
from dw4introduxi.dimension_table_session, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_page, dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_session.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_session
and dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
and page_id like '%info%'
```

```

intersect
select distinct dw4introduxi.dimension_table_session.id
from dw4introduxi.dimension_table_session, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_page, dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_session.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_session
and dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
and page_id like '%encomendar%'

```

Chart 9: Page request distribution over the observation period

```

select dw4introduxi.dimension_table_date.date_id,
count(dw4introduxi.dimension_table_page.page_id)
from dw4introduxi.dimension_table_date, dw4introduxi.dimension_table_page,
dw4introduxi.fact_table_usage
where dw4introduxi.dimension_table_date.dimension_key=
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_page.dimension_key=
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
Group by dw4introduxi.dimension_table_date.date_id
order by dw4introduxi.dimension_table_date.date_id asc

```

Chart 10: Number of sessions over the observation period

```

select dw4introduxi.dimension_table_date.date_id, count(distinct
dw4introduxi.dimension_table_session.id)
from dw4introduxi.dimension_table_date, dw4introduxi.dimension_table_session,
dw4introduxi.fact_table_usage
where dw4introduxi.dimension_table_session.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_session
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
Group by dw4introduxi.dimension_table_date.date_id
order by dw4introduxi.dimension_table_date.date_id asc

```

Chart 11: Page request distribution over the week for the observation period.

```

select dw4introduxi.dimension_table_date.week_day, count
(dw4introduxi.dimension_table_page.id)
from dw4introduxi.fact_table_usage, dw4introduxi.dimension_table_page,
dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
group by dw4introduxi.dimension_table_date.week_day

```

Chart 12: Orders distribution over the week for the observation period.

```

select dw4introduxi.dimension_table_date.week_day, count
(dw4introduxi.dimension_table_page.id)
from dw4introduxi.fact_table_usage, dw4introduxi.dimension_table_page,
dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
and page_id like '%encomendar%'
group by dw4introduxi.dimension_table_date.week_day

```

Chart 13: Page request distribution over the day for the observation period

```

select dw4introduxi.dimension_table_time.hour, count
(dw4introduxi.dimension_table_page.id)
from dw4introduxi.dimension_table_time, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_page, dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_time.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_time
and dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
group by dw4introduxi.dimension_table_time.hour
order by dw4introduxi.dimension_table_time.hour asc

```

Chart 14: Orders distribution over the day for the observation period

```
select dw4introduxi.dimension_table_time.hour, count
(dw4introduxi.dimension_table_page.id)
from dw4introduxi.dimension_table_time, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_page, dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_time.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_time
and dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
and page_id like '%encomendar%'
group by dw4introduxi.dimension_table_time.hour
order by dw4introduxi.dimension_table_time.hour asc
```

Chart 15: Top ten clickers.

```
select dw4introduxi.dimension_table_user.identification, count
(dw4introduxi.dimension_table_page.page_id) as p
from dw4introduxi.dimension_table_user, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_page, dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_user.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_user
and dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
```

```

and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-31'
group by dw4introduxi.dimension_table_user.identification
order by p desc

```

Chart 16: Top ten buyers.

```

select dw4introduxi.dimension_table_user.identification, count
(dw4introduxi.dimension_table_page.page_id) as p
from dw4introduxi.dimension_table_user, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_page, dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_user.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_user
and dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-31'
and page_id like '%encomendar%'
group by dw4introduxi.dimension_table_user.identification
order by p desc

```

Number of sessions containing a certain action page

```

select count (distinct dw4introduxi.dimension_table_session.session_id)
from dw4introduxi.dimension_table_session, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_page, dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_session.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_session

```

```

and dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
AND page_id like '%PORT%'
AND page_id LIKE '%info%'

```

Number of active sessions containing a certain action page and the target page

```

select dw4introduxi.dimension_table_session.id
from dw4introduxi.dimension_table_session, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_page, dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_session.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_session
and dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
and page_id like '%PORT%'
and page_id like '%info%'
intersect
select dw4introduxi.dimension_table_session.id
from dw4introduxi.dimension_table_session, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_page, dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_session.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_session
and dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page

```

```

and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
and page_id like '%PORT%'
and page_id like '%comprar%'
intersect
select dw4introduxi.dimension_table_session.id
from dw4introduxi.dimension_table_session, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_page, dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_session.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_session
and dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
and dw4introduxi.dimension_table_page.page_id like '%encomendar%'

```

Number of sessions containing a certain action page considering only customer sessions

```

select dw4introduxi.dimension_table_session.id
from dw4introduxi.dimension_table_session, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_page, dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_session.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_session
and dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date

```

```

and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
and page_id like '%PORT%'
and page_id like '%info%'
intersect
select dw4introduxi.dimension_table_session.id
from dw4introduxi.dimension_table_session, dw4introduxi.fact_table_usage,
dw4introduxi.dimension_table_page, dw4introduxi.dimension_table_date
where dw4introduxi.dimension_table_session.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_session
and dw4introduxi.dimension_table_page.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_page
and dw4introduxi.dimension_table_date.dimension_key =
dw4introduxi.fact_table_usage.dimension_table_date
and dw4introduxi.dimension_table_date.date_id BETWEEN '2008-01-01' and '2008-01-
31'
and dw4introduxi.dimension_table_page.page_id like '%encomendar%'

```

Appendix 4

Log Results

Products' Family ACESSORIOS	Reference Code	Contact efficiency	Relative Contact efficiency	Conversion efficiency
CAIXAS E BOLSAS	ACE	0,68%	1,07%	3,94%
ADAPTADORES	ADP	1,14%	1,79%	4,11%
ARQUIVADORES	ARQ	0,18%	0,29%	9,49%
BATERIAS	BAT	0,89%	1,39%	6,49%
CABOS	CB	1,56%	2,44%	5,77%
CAIXA EXT DISCO	DISCO	0,96%	1,50%	5,04%
DATA SWITCH	DATASWITCH	0,22%	0,35%	5,39%
FICHAS	FICHAS	0,04%	0,06%	17,24%
MALA PORTATIL	MALAP	1,11%	1,74%	4,35%
MATERIAL DIVERSO	DIV	0,21%	0,33%	3,16%
MICROFONES	MICROS	0,40%	0,63%	5,02%
MP3/MP4	MP3	1,51%	2,36%	4,64%
TAPETES	TAP	0,08%	0,13%	11,11%
TOMADAS	TOMADA	0,21%	0,34%	4,38%

Products' Family COMPONENTES	Reference Code	Contact efficiency	Relative Contact efficiency	Conversion efficiency
BAREBONE	BARE	0,74%	1,15%	0,18%
CAIXAS	CX	3,20%	5,01%	3,65%
CONTROLADORAS	CTRL	0,28%	0,43%	5,31%
COOLERS	COOLER	0,44%	0,69%	3,98%
CPU	CPU	1,34%	2,09%	4,02%
DISCOS	HD	1,41%	2,21%	6,27%
DRIVES DISQUETE	DRIVEDSK	0,16%	0,24%	3,45%
DRIVES DIVERSAS	DRIVEDVS	0,00%	0,00%	0,00%
LEITOR DVD E HD	DVDR	1,22%	1,91%	7,37%
FONTES	PW	0,75%	1,18%	5,36%
LEITOR DE CARTOES	CARDREADER	0,52%	0,82%	6,96%
MEMORIAS	RAM	1,86%	2,92%	5,69%
MOTHERBOARDS	MBA	0,86%	1,35%	2,34%
MOTHERBOARDS	MBI	3,49%	5,46%	5,69%
PLACAS GRAFICAS	PLGRA	3,55%	5,55%	3,56%
PLACAS REDE	PLR	0,26%	0,41%	4,06%
PLACAS SOM	PLSOM	0,06%	0,09%	9,52%
PLACAS VIDEO	PLVD	0,83%	1,29%	2,92%

Products' Family COMPUTADOR	Reference Code	Contact efficiency	Relative Contact efficiency	Conversion efficiency
COMPUTADORES	COMPDIV	0,12%	0,18%	0,00%
COMPUTADORES	COMPHP	0,29%	0,46%	1,83%
GOLDNET	GOLD	1,55%	2,42%	1,13%
GOLDNET	PC	1,69%	2,64%	1,11%
PORTATIL	PORT	4,35%	6,80%	0,83%
SERVIDORES	SERVIDORES	0,26%	0,41%	1,52%

Products' Family CONSUMIVEIS	Reference Code	Contact efficiency	Relative Contact efficiency	Conversion efficiency
CANETAS CDRPEN	CANETACDRPEN	0,01%	0,01%	0,00%
CASSETES	CASSETTE	0,01%	0,02%	12,50%
CD	CD	0,37%	0,58%	4,32%
COMPATIVEIS	COMPATIVEL	0,12%	0,19%	12,09%
DISQUETES	DISQUETES	0,02%	0,03%	0,00%
DVD	DVD	0,51%	0,80%	4,46%
PAPEL	PAPEL	0,03%	0,05%	4,17%
TINTEIROS	TINT	0,64%	1,01%	5,63%
TONNER	TONNER	0,53%	0,83%	1,77%

Products' Family IMAGEM	Reference Code	Contact efficiency	Relative Contact efficiency	Conversion efficiency
CAMARAS DIGITAIS	CAMDIG	0,94%	1,47%	2,01%
CAMARAS VIDEO	CV	0,21%	0,33%	0,64%
CARTOES MEMORIA	CARTMEM	0,84%	1,31%	6,25%
DVD SALA	DVDSALA	0,18%	0,28%	1,50%
DVDPORT	DVDPORT	0,07%	0,11%	1,96%
DVDRW SALA	DVDRWSALA	0,04%	0,06%	0,00%
PROJECTORES	PROJ	0,28%	0,43%	0,97%

Products' Family PERIFERICOS	Reference Code	Contact efficiency	Relative Contact efficiency	Conversion efficiency
BUNDLES	BUNDLE	1,56%	2,44%	0,26%
COLUNAS	COL	1,05%	1,64%	4,61%
DISCOS EXTERNOS	HDEXT	1,72%	2,69%	4,62%
DISCOS FLASH USB	FLASHU	1,17%	1,84%	6,98%
IMPRESSORAS	IMP	1,76%	2,75%	3,67%
JOYSTICKS	JOYSTICKS	0,37%	0,59%	5,38%
MULTIMEDIA	MM	1,06%	1,67%	1,39%
PLASMA/LCDTV	MONTFT	3,06%	4,79%	2,28%
POS	POS	0,69%	1,08%	4,84%
TECLADO	TEC	1,45%	2,27%	6,11%
RATOS	RATO	1,46%	2,28%	9,12%
UPS	UPS	0,66%	1,03%	5,08%
WEBCAM	WEBCAMC	0,82%	1,28%	7,57%

Products' Family REDES \ COMUNICAÇÃO	Reference Code	Contact efficiency	Relative Contact efficiency	Conversion efficiency
BLUETOOTH	BLTH	0,29%	0,46%	3,20%
FAX	FAX	0,07%	0,11%	1,96%
GPS	GPS	0,93%	1,46%	1,58%
HUBSWITCH	HUBS	0,55%	0,86%	4,67%
MODEMS ANALOGICOS	MODEMAN	0,04%	0,07%	0,00%
PDA	PDA	0,94%	1,47%	0,29%
REDE DIVERSOS	HUBDIV	0,31%	0,48%	3,07%
ROUTER	ROUTER	0,55%	0,85%	4,19%
TELEFONE	TELEF	0,91%	1,43%	2,06%
WIRELESS	WIRELESS	1,98%	3,10%	6,57%

Products' Family SOFTWARE	Reference Code	Contact efficiency	Relative Contact efficiency	Conversion efficiency
ANTIVIRUS	ANTV	0,17%	0,27%	3,94%
SOFTWARE	SWS	0,55%	0,86%	4,63%
SOFTWARE	SWO	0,52%	0,82%	3,87%
SOFTWARE	SWD	0,10%	0,16%	0,00%

Products' Family	Product type	Reference Code	Contact efficiency	Relative Contact efficiency	Conversion efficiency	AVG PRICE	Valor potencial de cada sessão que passa pela página A
CAIXAS E BOLSAS	ACESSORIOS	ACE	0,68%	1,07%	3,94%	25,36 €	0,05 €
ADAPTADORES	ACESSORIOS	ADP	1,14%	1,79%	4,11%	13,10 €	0,02 €
ARQUIVADORES	ACESSORIOS	ARQ	0,18%	0,29%	9,49%	12,34 €	0,09 €
BATERIAS	ACESSORIOS	BAT	0,89%	1,39%	6,49%	12,03 €	0,02 €
CABOS	ACESSORIOS	CB	1,56%	2,44%	5,77%	5,52 €	0,00 €
CAIXA EXT DISCO	ACESSORIOS	DISCO	0,96%	1,50%	5,04%	16,98 €	0,02 €
DATA SWITCH	ACESSORIOS	DATASWITCH	0,22%	0,35%	5,39%	22,75 €	0,14 €
FICHAS	ACESSORIOS	FICHAS	0,04%	0,06%	17,24%	2,71 €	0,09 €
MALA PORTATIL	ACESSORIOS	MALAP	1,11%	1,74%	4,35%	12,68 €	0,02 €
MATERIAL DIVERSO	ACESSORIOS	DIV	0,21%	0,33%	3,16%	5,14 €	0,03 €
MICROFONES	ACESSORIOS	MICROS	0,40%	0,63%	5,02%	4,04 €	0,01 €
MP3/MP4	ACESSORIOS	MP3	1,51%	2,36%	4,64%	25,02 €	0,02 €
TAPETES	ACESSORIOS	TAP	0,08%	0,13%	11,11%	2,68 €	0,04 €
TOMADAS	ACESSORIOS	TOMADA	0,21%	0,34%	4,38%	7,61 €	0,05 €
BAREBONE	COMPONENTES	BARE	0,74%	1,15%	0,18%	121,49 €	0,22 €
CAIXAS	COMPONENTES	CX	3,20%	5,01%	3,65%	25,36 €	0,01 €
CONTROLADORAS	COMPONENTES	CTRL	0,28%	0,43%	5,31%	11,71 €	0,06 €
COOLERS	COMPONENTES	COOLER	0,44%	0,69%	3,98%	4,17 €	0,01 €
CPU	COMPONENTES	CPU	1,34%	2,09%	4,02%	76,60 €	0,08 €
DISCOS	COMPONENTES	HD	1,41%	2,21%	6,27%	46,31 €	0,04 €
DRIVES DISQUETE	COMPONENTES	DRIVEDSK	0,16%	0,24%	3,45%	4,41 €	0,04 €
DRIVES DIVERSAS	COMPONENTES	DRIVEDVS	0,00%	0,00%	0,00%	4,41 €	
LEITOR DVD E HD	COMPONENTES	DVDR	1,22%	1,91%	7,37%	215,40 €	0,24 €
FONTES	COMPONENTES	PW	0,75%	1,18%	5,36%	15,76 €	0,03 €
LEITOR DE CARTOES	COMPONENTES	CARDREADER	0,52%	0,82%	6,96%	5,49 €	0,01 €
MEMORIAS	COMPONENTES	RAM	1,86%	2,92%	5,69%	20,99 €	0,02 €
MOTHERBOARDS	COMPONENTES	MBA	0,86%	1,35%	2,34%	44,44 €	0,07 €
MOTHERBOARDS	COMPONENTES	MBI	3,49%	5,46%	5,69%	50,22 €	0,02 €
PLACAS GRAFICAS	COMPONENTES	PLGRA	3,55%	5,55%	3,56%	76,49 €	0,03 €
PLACAS REDE	COMPONENTES	PLR	0,26%	0,41%	4,06%	19,63 €	0,10 €
PLACAS SOM	COMPONENTES	PLSOM	0,06%	0,09%	9,52%	7,64 €	0,18 €
PLACAS VIDEO	COMPONENTES	PLVD	0,83%	1,29%	2,92%	34,59 €	0,06 €
COMPUTADORES	COMPUTADOR	COMPDIV	0,12%	0,18%	0,00%	499,00 €	5,74 €
COMPUTADORES	COMPUTADOR	COMPHP	0,29%	0,46%	1,83%	570,43 €	2,60 €
GOLDNET	COMPUTADOR	GOLD	1,55%	2,42%	1,13%	341,42 €	0,30 €
GOLDNET	COMPUTADOR	PC	1,69%	2,64%	1,11%	341,42 €	0,27 €
PORTATIL	COMPUTADOR	PORT	4,35%	6,80%	0,83%	710,36 €	0,22 €
SERVIDORES	COMPUTADOR	SERVIDORES	0,26%	0,41%	1,52%	299,03 €	1,52 €
CANETAS CDRPEN	CONSUMIVEIS	CANETACDRPEN	0,01%	0,01%	0,00%	2,91 €	0,42 €
CASSETES	CONSUMIVEIS	CASSETE	0,01%	0,02%	12,50%	8,07 €	1,01 €
CD	CONSUMIVEIS	CD	0,37%	0,58%	4,32%	4,55 €	0,02 €
COMPATIVES	CONSUMIVEIS	COMPATIVEL	0,12%	0,19%	12,09%	1,62 €	0,02 €
DISQUETES	CONSUMIVEIS	DISQUETES	0,02%	0,03%	0,00%	1,19 €	0,08 €
DVD	CONSUMIVEIS	DVD	0,51%	0,80%	4,46%	5,37 €	0,01 €
PAPEL	CONSUMIVEIS	PAPEL	0,03%	0,05%	4,17%	8,09 €	0,34 €
TINTEIROS	CONSUMIVEIS	TINT	0,64%	1,01%	5,63%	16,09 €	0,03 €
TONNER	CONSUMIVEIS	TONNER	0,53%	0,83%	1,77%	77,49 €	0,20 €
CAMARAS DIGITAIS	IMAGEM	CAMDIG	0,94%	1,47%	2,01%	139,40 €	0,20 €
CAMARAS VIDEO	IMAGEM	CV	0,21%	0,33%	0,64%	242,93 €	1,55 €
CARTOES MEMORIA	IMAGEM	CARTMEM	0,84%	1,31%	6,25%	10,02 €	0,02 €
DVD SALA	IMAGEM	DVDSALA	0,18%	0,28%	1,50%	122,33 €	0,92 €
DVDPORT	IMAGEM	DVDPORT	0,07%	0,11%	1,96%	121,07 €	2,37 €
DVDRW SALA	IMAGEM	DVDRWSALA	0,04%	0,06%	0,00%	171,36 €	6,12 €
PROJECTORES	IMAGEM	PROJ	0,28%	0,43%	0,97%	543,70 €	2,64 €
BUNDLES	PERIFERICOS	BUNDLE	1,56%	2,44%	0,26%	404,80 €	0,35 €
COLUNAS	PERIFERICOS	COL	1,05%	1,64%	4,61%	15,84 €	0,02 €
DISCOS EXTERNOS	PERIFERICOS	HDEXT	1,72%	2,69%	4,62%	72,08 €	0,06 €
DISCOS FLASH USB	PERIFERICOS	FLASHU	1,17%	1,84%	6,98%	9,18 €	0,01 €
IMPRESSORAS	PERIFERICOS	IMP	1,76%	2,75%	3,67%	73,67 €	0,06 €
JOYSTICKS	PERIFERICOS	JOYSTICKS	0,37%	0,59%	5,38%	11,93 €	0,04 €
MULTIMEDIA	PERIFERICOS	MM	1,06%	1,67%	1,39%	91,07 €	0,11 €
PLASMA/LCDTV	PERIFERICOS	MONTFT	3,06%	4,79%	2,28%	555,07 €	0,24 €
POS	PERIFERICOS	POS	0,69%	1,08%	4,84%	57,68 €	0,11 €
TECLADO	PERIFERICOS	TEC	1,45%	2,27%	6,11%	9,65 €	0,01 €
RATOS	PERIFERICOS	RATO	1,46%	2,28%	9,12%	7,59 €	0,01 €
UPS	PERIFERICOS	UPS	0,66%	1,03%	5,08%	62,79 €	0,13 €
WEBCAM	PERIFERICOS	WEBCAMC	0,82%	1,28%	7,57%	16,87 €	0,03 €
BLUETOOTH	REDES/COMUNICAÇÃO	BLTH	0,29%	0,46%	3,20%	11,00 €	0,05 €
FAX	REDES/COMUNICAÇÃO	FAX	0,07%	0,11%	1,96%	83,68 €	1,64 €
GPS	REDES/COMUNICAÇÃO	GPS	0,93%	1,46%	1,58%	75,81 €	0,11 €
HUB/SWITCH	REDES/COMUNICAÇÃO	HUBS	0,55%	0,86%	4,67%	12,92 €	0,03 €
MODEMS ANALOGICOS	REDES/COMUNICAÇÃO	MODEMAN	0,04%	0,07%	0,00%	22,10 €	0,67 €
PDA	REDES/COMUNICAÇÃO	PDA	0,94%	1,47%	0,29%	312,08 €	0,45 €
REDE DIVERSOS	REDES/COMUNICAÇÃO	HUBDIV	0,31%	0,48%	3,07%	51,64 €	0,23 €
ROUTER	REDES/COMUNICAÇÃO	ROUTER	0,55%	0,85%	4,19%	34,77 €	0,09 €
TELEFONE	REDES/COMUNICAÇÃO	TELEF	0,91%	1,43%	2,06%	31,34 €	0,05 €
WIRELESS	REDES/COMUNICAÇÃO	WIRELESS	1,98%	3,10%	6,57%	35,49 €	0,02 €
ANTIVIRUS	SOFTWARE	ANTV	0,17%	0,27%	3,94%	33,82 €	0,27 €
SOFTWARE	SOFTWARE	SWS	0,55%	0,86%	4,63%	81,15 €	0,20 €
SOFTWARE	SOFTWARE	SWO	0,52%	0,82%	3,87%	81,15 €	0,21 €
SOFTWARE	SOFTWARE	SWD	0,10%	0,16%	0,00%	81,15 €	1,04 €

Customer Session Log Results

Products' Family ACESSORIOS	Reference Code	Relative Contact efficiency	Conversion efficiency
CAIXAS E BOLSAS	ACE	1,87%	46,51%
ADAPTADORES	ADP	4,39%	34,65%
ARQUIVADORES	ARQ	0,96%	59,09%
BATERIAS	BAT	3,21%	58,11%
CABOS	CB	4,95%	58,77%
CAIXA EXT DISCO	DISCO	3,13%	50,00%
DATA SWITCH	DATASWITCH	0,70%	56,25%
FICHAS	FICHAS	0,22%	100,00%
MALA PORTATIL	MALAP	2,78%	56,25%
MATERIAL DIVERSO	DIV	0,65%	33,33%
MICROFONES	MICROS	0,91%	71,43%
MP3/MP4	MP3	3,74%	60,47%
TAPETES	TAP	0,56%	53,85%
TOMADAS	TOMADA	0,91%	33,33%

Products' Family COMPONENTES	Reference Code	Relative Contact efficiency	Conversion efficiency
BAREBONE	BARE	1,22%	3,57%
CAIXAS	CX	7,30%	51,79%
CONTROLADORAS	CTRL	0,96%	50,00%
COOLERS	COOLER	1,04%	54,17%
CPU	CPU	3,39%	51,28%
DISCOS	HD	4,60%	62,26%
DRIVES DISQUETE	DRIVEDSK	0,35%	50,00%
DRIVES DIVERSAS	DRIVEDVS	0,00%	0,00%
LEITOR DVD E HD	DVDR	4,30%	67,68%
FONTES	PW	2,17%	60,00%
LEITOR DE CARTOES	CARDREADER	2,09%	56,25%
MEMORIAS	RAM	5,39%	63,71%
MOTHERBOARDS	MBA	2,48%	26,32%
MOTHERBOARDS	MBI	10,17%	63,25%
PLACAS GRAFICAS	PLGRA	6,47%	63,09%
PLACAS REDE	PLR	0,70%	50,00%
PLACAS SOM	PLSOM	0,26%	66,67%
PLACAS VIDEO	PLVD	1,61%	48,65%

Products' Family COMPUTADOR	Reference Code	Relative Contact efficiency	Conversion efficiency
COMPUTADORES	COMPDIV	0,17%	0,00%
COMPUTADORES	COMPHP	0,48%	36,36%
GOLDNET	GOLD	2,17%	26,00%
GOLDNET	PC	2,09%	29,17%
PORTATIL	PORT	3,82%	30,68%
SERVIDORES	SERVIDORES	0,56%	23,08%

Products' Family CONSUMIVEIS	Reference Code	Relative Contact efficiency	Conversion efficiency
CANETAS CDRPEN	CANETACDRPEN	0,00%	n/d
CASSETES	CASSETE	0,04%	100,00%
CD	CD	0,70%	75,00%
COMPATIVEIS	COMPATIVEL	0,52%	91,67%
DISQUETES	DISQUETES	0,04%	0,00%
DVD	DVD	1,35%	54,84%
PAPEL	PAPEL	0,04%	100,00%
TINTEIROS	TINT	1,82%	64,29%
TONNER	TONNER	0,48%	63,64%

Products' Family IMAGEM	Reference Code	Relative Contact efficiency	Conversion efficiency
CAMARAS DIGITAIS	CAMDIG	1,65%	36,84%
CAMARAS VIDEO	CV	0,30%	14,29%
CARTOES MEMORIA	CARTMEM	2,13%	79,59%
DVD SALA	DVDSALA	0,48%	18,18%
DVDPORT	DVDPORT	0,13%	33,33%
DVDRW SALA	DVDRWSALA	0,00%	n/d
PROJECTORES	PROJ	0,43%	20,00%

Products' Family PERIFERICOS	Reference Code	Relative Contact efficiency	Conversion efficiency
BUNDLES	BUNDLE	2,56%	5,08%
COLUNAS	COL	3,08%	50,70%
DISCOS EXTERNOS	HDEXT	4,52%	56,73%
DISCOS FLASH USB	FLASHU	4,13%	64,21%
IMPRESSORAS	IMP	3,95%	52,75%
JOYSTICKS	JOYSTICKS	1,00%	65,22%
MULTIMEDIA	MM	2,00%	23,91%
PLASMA/LCDTV	MONTFT	5,26%	42,98%
POS	POS	2,00%	54,35%
TECLADO	TEC	4,65%	61,68%
RATOS	RATO	5,82%	73,88%
UPS	UPS	1,82%	59,52%
WEBCAM	WEBCAMC	3,04%	65,71%

Products' Family REDES \ COMUNICAÇÃO	Reference Code	Relative Contact efficiency	Conversion efficiency
BLUETOOTH	BLTH	0,61%	50,00%
FAX	FAX	0,22%	20,00%
GPS	GPS	1,65%	28,95%
HUBSWITCH	HUBS	1,56%	52,78%
MODEMS ANALOGICOS	MODEMAN	0,04%	0,00%
PDA	PDA	1,17%	7,41%
REDE DIVERSOS	HUBDIV	0,65%	46,67%
ROUTER	ROUTER	1,69%	43,59%
TELEFONE	TELEF	1,35%	45,16%
WIRELESS	WIRELESS	6,73%	62,58%

Products' Family SOFTWARE	Reference Code	Relative Contact efficiency	Conversion efficiency
ANTIVIRUS	ANTV	0,30%	71,43%
SOFTWARE	SWS	1,61%	51,35%
SOFTWARE	SWO	1,26%	51,72%
SOFTWARE	SWD	0,22%	0,00%

Products' Family	Product type	Reference Code	Relative Contact efficiency	Conversion efficiency	AVG PRICE	Valor potencial de cada sessão que passa pela página A
CAIXAS E BOLSAS	ACESSORIOS	ACE	1,87%	46,51%	25,36 €	0,59 €
ADAPTADORES	ACESSORIOS	ADP	4,39%	34,65%	13,10 €	0,13 €
ARQUIVADORES	ACESSORIOS	ARQ	0,96%	59,09%	12,34 €	0,56 €
BATERIAS	ACESSORIOS	BAT	3,21%	58,11%	12,03 €	0,16 €
CABOS	ACESSORIOS	CB	4,95%	58,77%	5,52 €	0,05 €
CAIXA EXT DISCO	ACESSORIOS	DISCO	3,13%	50,00%	16,98 €	0,24 €
DATA SWITCH	ACESSORIOS	DATASWITCH	0,70%	56,25%	22,75 €	1,42 €
FICHAS	ACESSORIOS	FICHAS	0,22%	100,00%	2,71 €	0,54 €
MALA PORTATIL	ACESSORIOS	MALAP	2,78%	56,25%	12,68 €	0,20 €
MATERIAL DIVERSO	ACESSORIOS	DIV	0,65%	33,33%	5,14 €	0,34 €
MICROFONES	ACESSORIOS	MICROS	0,91%	71,43%	4,04 €	0,19 €
MP3/MP4	ACESSORIOS	MP3	3,74%	60,47%	25,02 €	0,29 €
TAPETES	ACESSORIOS	TAP	0,56%	53,85%	2,68 €	0,21 €
TOMADAS	ACESSORIOS	TOMADA	0,91%	33,33%	7,61 €	0,36 €
BAREBONE	COMPONENTES	BARE	1,22%	3,57%	121,49 €	4,34 €
CAIXAS	COMPONENTES	CX	7,30%	51,79%	25,36 €	0,15 €
CONTROLADORAS	COMPONENTES	CTRL	0,96%	50,00%	11,71 €	0,53 €
COOLERS	COMPONENTES	COOLER	1,04%	54,17%	4,17 €	0,17 €
CPU	COMPONENTES	CPU	3,39%	51,28%	76,60 €	0,98 €
DISCOS	COMPONENTES	HD	4,60%	62,26%	46,31 €	0,44 €
DRIVES DISQUETE	COMPONENTES	DRIVEDSK	0,35%	50,00%	4,41 €	0,55 €
DRIVES DIVERSAS	COMPONENTES	DRIVEDVS	0,00%	0,00%	4,41 €	
LEITOR DVD E HD	COMPONENTES	DVDR	4,30%	67,68%	215,40 €	2,18 €
FONTES	COMPONENTES	PW	2,17%	60,00%	15,76 €	0,32 €
LEITOR DE CARTOES	COMPONENTES	CARDREADER	2,09%	56,25%	5,49 €	0,11 €
MEMORIAS	COMPONENTES	RAM	5,39%	63,71%	20,99 €	0,17 €
MOTHERBOARDS	COMPONENTES	MBA	2,48%	26,32%	44,44 €	0,78 €
MOTHERBOARDS	COMPONENTES	MBI	10,17%	63,25%	50,22 €	0,21 €
PLACAS GRAFICAS	COMPONENTES	PLGRA	6,47%	63,09%	76,49 €	0,51 €
PLACAS REDE	COMPONENTES	PLR	0,70%	50,00%	19,63 €	1,23 €
PLACAS SOM	COMPONENTES	PLSOM	0,26%	66,67%	7,64 €	1,27 €
PLACAS VIDEO	COMPONENTES	PLVD	1,61%	48,65%	34,59 €	0,93 €
COMPUTADORES	COMPUTADOR	COMPDIV	0,17%	0,00%	499,00 €	124,75 €
COMPUTADORES	COMPUTADOR	COMPHP	0,48%	36,36%	570,43 €	51,86 €
GOLDNET	COMPUTADOR	GOLD	2,17%	26,00%	341,42 €	6,83 €
GOLDNET	COMPUTADOR	PC	2,09%	29,17%	341,42 €	7,11 €
PORTATIL	COMPUTADOR	PORT	3,82%	30,68%	710,36 €	8,07 €
SERVIDORES	COMPUTADOR	SERVIDORES	0,56%	23,08%	299,03 €	23,00 €
CANETAS CDRPEN	CONSUMIVEIS	CANETACDRPEN	0,00%	n/d	2,91 €	
CASSETES	CONSUMIVEIS	CASSETE	0,04%	100,00%	8,07 €	8,07 €
CD	CONSUMIVEIS	CD	0,70%	75,00%	4,55 €	0,28 €
COMPATIVEIS	CONSUMIVEIS	COMPATIVEL	0,52%	91,67%	1,62 €	0,13 €
DISQUETES	CONSUMIVEIS	DISQUETES	0,04%	0,00%	1,19 €	1,19 €
DVD	CONSUMIVEIS	DVD	1,35%	54,84%	5,37 €	0,17 €
PAPEL	CONSUMIVEIS	PAPEL	0,04%	100,00%	8,09 €	8,09 €
TINTEIROS	CONSUMIVEIS	TINT	1,82%	64,29%	16,09 €	0,38 €
TONNER	CONSUMIVEIS	TONNER	0,48%	63,64%	77,49 €	7,04 €
CAMARAS DIGITAIS	IMAGEM	CAMDIG	1,65%	36,84%	139,40 €	3,67 €
CAMARAS VIDEO	IMAGEM	CV	0,30%	14,29%	242,93 €	34,70 €
CARTOES MEMORIA	IMAGEM	CARTMEM	2,13%	79,59%	10,02 €	0,20 €
DVD SALA	IMAGEM	DVDSALA	0,48%	18,18%	122,33 €	11,12 €
DVDPORT	IMAGEM	DVDPORT	0,13%	33,33%	121,07 €	40,36 €
DVDRW SALA	IMAGEM	DVDRWSALA	0,00%	n/d	171,36 €	
PROJECTORES	IMAGEM	PROJ	0,43%	20,00%	543,70 €	54,37 €
BUNDLES	PERIFERICOS	BUNDLE	2,56%	5,08%	404,80 €	6,86 €
COLUNAS	PERIFERICOS	COL	3,08%	50,70%	15,84 €	0,22 €
DISCOS EXTERNOS	PERIFERICOS	HDEXT	4,52%	56,73%	72,08 €	0,69 €
DISCOS FLASH USB	PERIFERICOS	FLASHU	4,13%	64,21%	9,18 €	0,10 €
IMPRESSORAS	PERIFERICOS	IMP	3,95%	52,75%	73,67 €	0,81 €
JOYSTICKS	PERIFERICOS	JOYSTICKS	1,00%	65,22%	11,93 €	0,52 €
MULTIMEDIA	PERIFERICOS	MM	2,00%	23,91%	91,07 €	1,98 €
PLASMA/LCDTV	PERIFERICOS	MONTFT	5,26%	42,98%	555,07 €	4,59 €
POS	PERIFERICOS	POS	2,00%	54,35%	57,68 €	1,25 €
TECLADO	PERIFERICOS	TEC	4,65%	61,68%	9,65 €	0,09 €
RATOS	PERIFERICOS	RATO	5,82%	73,88%	7,59 €	0,06 €
UPS	PERIFERICOS	UPS	1,82%	59,52%	62,79 €	1,50 €
WEBCAM	PERIFERICOS	WEBCAMC	3,04%	65,71%	16,87 €	0,24 €
BLUETOOTH	REDES/COMUNICAÇÃO	BLTH	0,61%	50,00%	11,00 €	0,79 €
FAX	REDES/COMUNICAÇÃO	FAX	0,22%	20,00%	83,68 €	16,74 €
GPS	REDES/COMUNICAÇÃO	GPS	1,65%	28,95%	75,81 €	2,00 €
HUB/SWITCH	REDES/COMUNICAÇÃO	HUBS	1,56%	52,78%	12,92 €	0,36 €
MODEMS ANALOGICOS	REDES/COMUNICAÇÃO	MODEMAN	0,04%	0,00%	22,10 €	22,10 €
PDA	REDES/COMUNICAÇÃO	PDA	1,17%	7,41%	312,08 €	11,56 €
REDE DIVERSOS	REDES/COMUNICAÇÃO	HUBDIV	0,65%	46,67%	51,64 €	3,44 €
ROUTER	REDES/COMUNICAÇÃO	ROUTER	1,69%	43,59%	34,77 €	0,89 €
TELEFONE	REDES/COMUNICAÇÃO	TELEF	1,35%	45,16%	31,34 €	1,01 €
WIRELESS	REDES/COMUNICAÇÃO	WIRELESS	6,73%	62,58%	35,49 €	0,23 €
ANTIVIRUS	SOFTWARE	ANTV	0,30%	71,43%	33,82 €	4,83 €
SOFTWARE	SOFTWARE	SWS	1,61%	51,35%	81,15 €	2,19 €
SOFTWARE	SOFTWARE	SWO	1,26%	51,72%	81,15 €	2,80 €
SOFTWARE	SOFTWARE	SWD	0,22%	0,00%	81,15 €	16,23 €