Using Web Mining

João Dias, João Miranda ,João Sousa, Henrique Fernandes, Olga Popova, Isabel Seruca

1) Universidade Portucalense, Porto, Portugal

joao.moreira.dias1@gmail.com

2) Universidade Portucalense, Porto, Portugal

3) Universidade Portucalense, Porto, Portugal

joao\_filipe\_sousa@hotmail.com

4) Universidade Portucalense, Porto, Portugal

[hspotted@gmail.com](mailto:hspotted@gmail.com)

5) Universidade Portucalense, Porto, Portugal

[olgapopova.y](mailto:sergiopascoal@gmail.com)@gmail.com

6) Universidade Portucalense, Porto, Portugal

[iseruca@upt.pt](mailto:iseruca@upt.pt)

# Introducing

To be successful in the business environment is a condition for improving the efficiency of information technology. It requires the ability to work competently with information, using an automated management system that provides centralized storage, management, and access to it.

The necessity automatic analysis of information from the Internet due to the high accessibility the continuously expanding huge amount of information, as well as the growing popularity of Web services among all users. The development of the Web into a global information infrastructure has allowed normal users to not only be consumers of information, but its creators and distributors. In this context, to address effectively the problems of searching, structuring and analyzing are mainly chaotically organized information in the intended a new direction in the methodology of data analysis - Web Mining. Web mining can be generally defined as the discovery and analysis of useful information from the World Wide Web. It describes the automatic search for information resources accessible online that allows gain a competitive advantage that helps to collect, store and monitor and analyze the data and transform them into useful business information.In this article we show a practical case of application of these technologies in practice.

The work reported here in was developed in the context of an Erasmus Intensive Programme (IP) Project, where several European higher education partner institutions were involved. The long-run objective of the IP WISDOM (Web Information System Data Organization Modeling) is to build an international curriculum in which the partners can subscribe. Each edition of the IP project is, therefore, a part of a long-term project to develop a European Curriculum for Web Mining studies.

The paper is structured as follows: Section 2 presents a description of the practical case in which the project was based and sets the objectives to fulfill during the project. Section 3 provides a description of the Key Performance Indicators (KPI’s) that were considered at the Web Mining. In Section 4 we present an outline of the way-of-working and present the major outcomes obtained with the project. Section 5 concludes with a summary of the project achievements and our directions for future work.

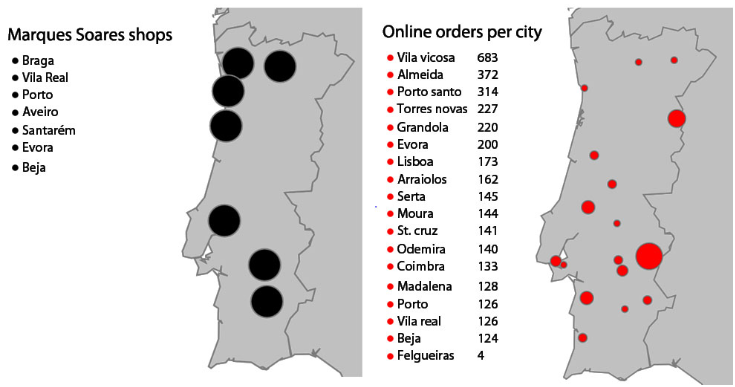
# Case Data: The goals and objectives of the study

Company Marques Soares, founded in 1960, mainly specializes in women's, men's and children's clothes. The company has a online shop and physical shops located in Portugal, in such cities as Porto, Braga, Aveiro, Beja, Santarém, Évora, and Vila Real. The company wants take advantage of Web Mining in order to be able to:

* winning new customers
* the opportunity for current and new customers use the website as a tool to purchase products
* provide an opportunity for customers to monitor their accounts
* development of an information channel for the company

To achieve this company wants to use a wide range of digital customer data and store this data in a secure environment, analyze and produce reports. Thus, it will be possible to identify customer preferences, to identify a core group of active buyers, and to identify the cause of purchases (such as seasonal discounts, and easy to use online shop).

Customers who buy through the Internet, and those who buy in the stores are different. Most customers shop Marques Soares are middle-aged women. The majority of purchases over the Internet to make more young people than those who go and buy clothes in the store. need to know how to better to attract a buyer a certain age category. It is also necessary not only to attract new customers, but try not to miss the old ones.



The main objective is study that analyzes the consumer behavior of buyers clothing, shoes and accessories online (E-business plataform) www.marquessoares.pt. For this considered such factors as gender, age, place of residence, the cost and frequency of shopping preferred by methods of payment and articles of clothing. In the research report presented the popularity ratings of certain purchases.

In this work we investigate how a firm can adapt its business strategy on the Internet and how capture a larger range of customers and thus attract more new customers by means of social networking.

The main aims and objectives of the work:

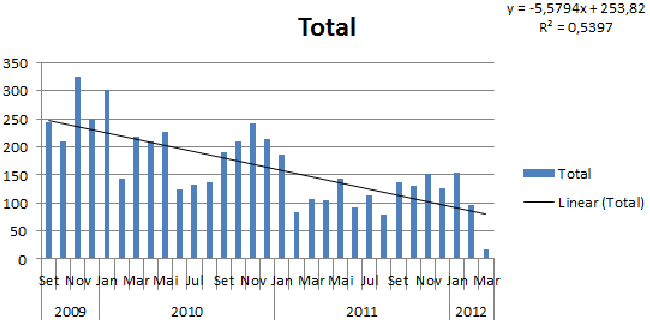
* Using Google Analityc to analyze Internet traffic
* Analyze the visitors which sites were still
* What are the key words have been used to search the site
* Evaluation of user behavior and building sales ranking
* Assess the scope and structure of online shop Marques Soares
* Determination of the influence of geographical factors on the behavior and preferences of consumers of clothing, footwear and accessories through the Internet - the place of residence, sex, age, preferred pricing.

# KPI

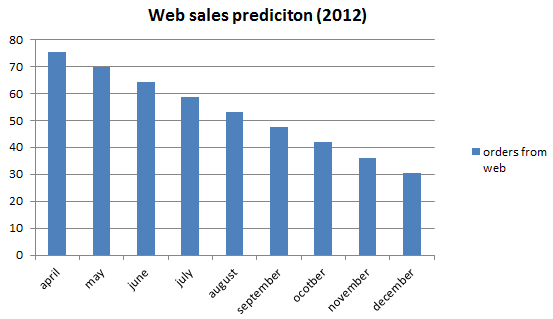
# Processing and analysis of data

## 4.1 Web Transaction

Graph of the amount of online orders on the site. It´s clear that it decreasing during time. This is not a good trend at all, because it is critical for everything we do on the website. We tracked the results out of excel, with table--‐mining tools. Marques Soares has to improve these numbers, to compete with their concurrent.

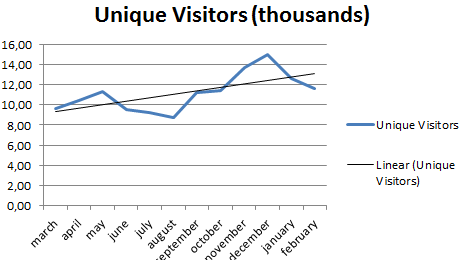


Out of Previous results, was made the Function “y= ‐5.5794x+253.82” out of. With this function we can make predictions of web sales in future. The only thing you have to do is replace the x of the number of months. This is based on the trend line of the table, so not accurate. Also based on the past so they can change.

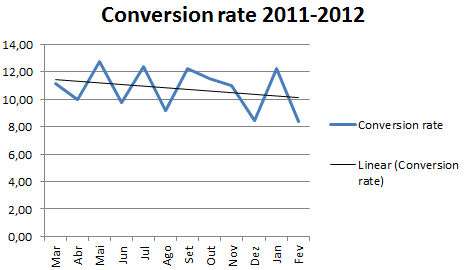


## 4.2 Trafic Conversion

Amount of unique people visiting the site is increasing. This is a good thing because more people will know something about the website. Especially the months in the winter are very important, probably due to the weather, because these are higher that in the summer. So they could do more promotion in the winter.

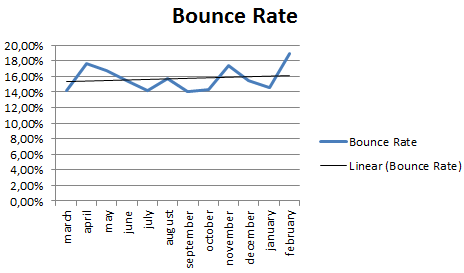


More customers the percentage of coversion doesn´t incrise. It’s was proved with conversion rate figure, the function was used is “orders divided by the unique visitors”.

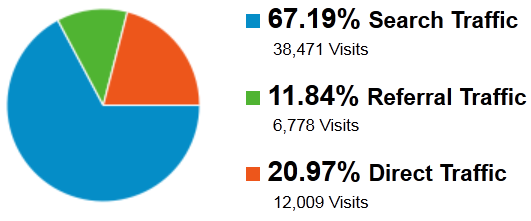


## 4.3 Web Statistics

Next picture represents the bounce rate that we researched in Google Analytics. The bounce rate is when surfers come on the site and immediately go away. This is bad because they wouldn’t by anything. We have to attract them more, so when they arrived at the home page. Even though it was accidental, they would still give a look.



The next chart represents the actual numbers of what the route of the visitors. The biggest part of the visitors went to the site via search engines. This is rather normal, because search engine are really popular. But necessary to improve the referral traffic, because this means that we are linked with a lot of sites.



Of the 67,19% that enter the site via search traffic, we want to know what the keywords were. And the results were pretty impressive, they just enter “Marques Soares”, this means that we have a really good brand name. So we suggest that we also go wide. If enter “Portuguese shop” in Portuguese language can see Marques Soares shining at the top.

Important analyze duration of a visit, because we want that the customers is as long as possible on the site. This is important because it results in bigger profits. A possible way to do this is via data mining, because when they recommend personal products customers are more likely to buy something.

# Conclusion

After completing this project, we believe that we have achieved good results when using Google Analityc, we hope that our results will benefit the company Marques Soares.  
 Unfortunately, in the during of work we were unable to find information about the real age of customers, because the family cards are used for purchases are valid for all family members, so the data on the average age category of customers could be distorted. With such imprecise data is dangerous to make any findings and conclusions to the long periods because of possible inaccuracies in the data.

# References