



# **ONLINE DESTINATION IMAGE OF PORTO: A CONSUMER BASED PERSPECTIVE**

**ANA RITA CARVALHO MOURA DE SOUSA**

**MASTER'S DISSERTATION IN MARKETING**

**ADVISOR: PROF. DOUTORA MARIA CATARINA DE ALMEIDA ROSEIRA**

**CO-ADVISOR PROFESSOR: DOUTOR JOÃO MANUEL PORTELA DA GAMA**

**2014**

## Short biographical presentation

Ana Rita Carvalho Moura de Sousa was born in 1985 in Porto, Portugal. In 2008 she completed her bachelor in Management at the Faculty of Economics of the University of Porto. In September 2013, she enrolled as a graduate student in the Master in Marketing at the University of Porto. In the second term of the first year of studies she had the opportunity to master her knowledge in the University of Carlos III, in Madrid, as an exchange student. Since 2010, she has been working in the retail sector.

## Abstract

Understand the tourist's image of a place is an important piece of information that tourism marketers should be aware of. Especially in the Web 2.0 era where tourists not only consume but also contribute with information. Additionally, studies suggest that User Generated Content (UGC) have influence in the purchase decision making process (O'Connor, 2008). Many studies have already explored online Destination Image but none have addressed Porto's region. This study analyses data collected from reviews that have been published online on popular blog website Travepod and travel website Tripadvisor. In the last one, the study focus on Attractions, Restaurants and Accommodation categories. The methodology applied is based on text mining tools and focuses on words frequency. Results highlights a visitors' association to Douro River and Port Wine. Results seem also to indicate a positive image of Porto, with good results in all four studied categories.

**Keywords:** Destination Image, Travel 2.0, Electronic Word Of Mouth, User Generated Content, Text mining, Travel Blogs, Tripadvisor, Porto.

# Index

<b>SHORT BIOGRAPHICAL PRESENTATION .....</b>	<b>III</b>
<b>ABSTRACT .....</b>	<b>IV</b>
<b>1. INTRODUCTION .....</b>	<b>1</b>
<b>2. DESTINATION IMAGE .....</b>	<b>3</b>
2.1. DESTINATION IMAGE FORMATION AND DIMENSIONS .....	6
<b>3. TRAVEL 2.0. ....</b>	<b>11</b>
3.1. ELECTRONIC WORD OF MOUTH (E-WOM) .....	14
3.2. USER GENERATED CONTENT (UGC) .....	14
3.2.1. WEBLOGS .....	17
<b>4. RESEARCH GOAL AND METHODOLOGY .....</b>	<b>20</b>
4.1. DATA COLLECTION .....	22
4.2. DATA ANALYSIS .....	23
<b>5. RESULTS.....</b>	<b>25</b>
5.1. BLOG.....	25
5.2. TRIPADVISOR .....	28
ATTRACTIONS.....	28
RESTAURANTS .....	31
ACCOMMODATION.....	33
<b>6. CONCLUSIONS .....</b>	<b>35</b>
6.1. CONTRIBUTIONS .....	37
6.2. LIMITATIONS AND FUTURE RECOMMENDATIONS .....	37
<b>7. BIBLIOGRAPHY .....</b>	<b>38</b>
<b>ANNEX 1 – RAPID MINER.....</b>	<b>43</b>
<b>ANNEX 2 - BLOG DATA .....</b>	<b>45</b>
<b>ANNEX 3 - TRIPADVISOR ATTRACTIONS DATA .....</b>	<b>49</b>

<b>ANNEX 4 - TRIPADVISOR RESTAURANTS DATA .....</b>	<b>53</b>
<b>ANNEX 5 - TRIPADVISOR ACCOMMODATION DATA .....</b>	<b>57</b>

## Figures index

Figure 1 - A general framework of Destination Image formation.....	6
Figure 2 - 7 Stages model theory of Destination Image formation .....	7
Figure 3 – Attributes to measure Destination Image.....	9
Figure 4 - Destination Image components.....	10
Figure 5 - RapidMiner process resume.....	24
Figure 6 – Destination Image components – adaptation to Porto.....	35

## Tables index

Table 1 – TravelPod - top bigrams by nationality .....	27
Table 2- Tripadvisor Attractions - top bigrams by nationality .....	30
Table 4 - TripAdvisor Restaurants - top bigrams by nationality .....	32
Table 5 - TripAdvisor Accommodation - top bigrams by nationality .....	34

## Illustrations index

Illustration 1 - Blog cloud.....	25
Illustration 2 - Attractions cloud.....	28
Illustration 3 - Restaurants cloud.....	31
Illustration 4 - Accommodation cloud.....	33

## Graphics index

Graphic 1 - Inspiration and travel planning sources.....	1
Graphic 2 - Word best describe Portugal .....	5
Graphic 3 - Number of individuals using internet, 2005-2014 (estimated).....	11
Graphic 4 - Degrees of trust per type of advertising .....	16

# Acronyms

GDP – Gross Domestic Product

e-WOM – Electronic Word of Mouth

ICT - Information and Communication Technologies

WTTC - World Travel and Tourism Council

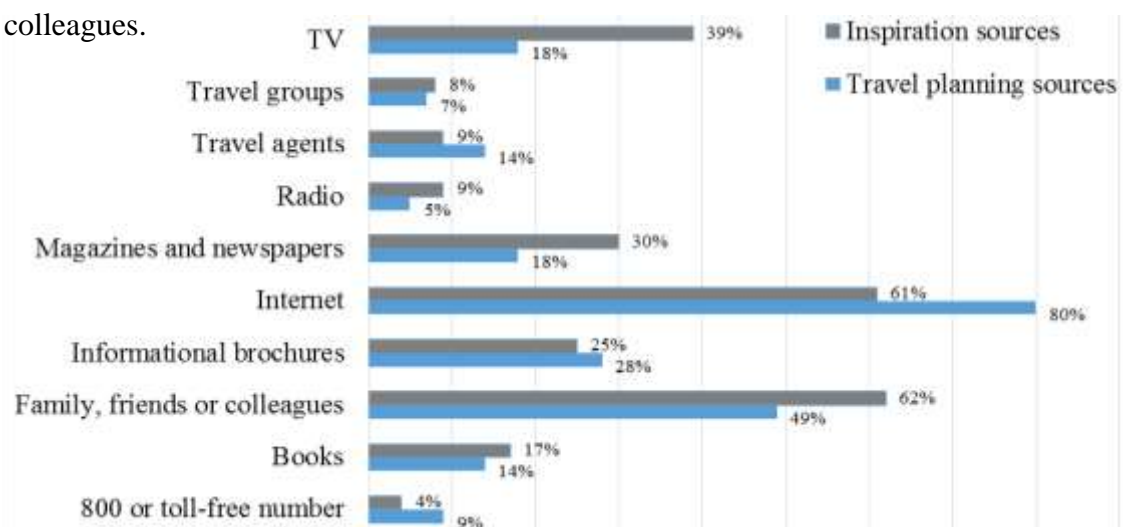
WOM - Word of Mouth

UGC – User Generated Content

# 1.Introduction

Destination Image has been raising the attention of social scientists. This concept comprises the emotions, feelings, reactions and effects that are triggered and perceived about a particular destination (Echtner & Ritchie, 1993). Its importance is twofold: it influences consumers choice on their tourism destination and afterwards, when evaluating the trip and future intentions (Reza, Samiei, Dini, & Yaghoubi, 2012). More recently, studies about Destination Image in the internet context are following the development of information and communication technologies (ICT). With the broad use of Internet it is possible to collect information and assess opinions and judgements of products images, services or places without coming into direct contact with consumers.

Online holidays reviews can influence individual travellers and therefore impact the overall competition of the tourism sector. This new reality has changed the market from a customer-centric to a customer-driven approach, where consumers not only follow the information but also share and contribute to it (Buhalis & Jun, 2011). Moreover, studies suggest that interpersonal influence and Word Of Mouth (WOM) are actually seen as the most important source of information when considering a purchase. Google (2013a) published a study (with Ipsos Media CT) that reinforces this idea. In the study, interviewees were questioned about the sources that typically inspire them to start thinking about a personal or leisure trip and the sources they often use to plan a trip (graphic 1). The majority of interviewees make use of the internet and family, friends or colleagues.



Graphic 1 - Inspiration and travel planning sources  
Source: Adapted from Ipsos MediaCT/Google Travel Study (2013)

In this current era, understanding how users perceive and assess online holiday reviews is not just relevant for practitioners but it is also challenging for academics (Papathanassis & Knolle, 2011).

In the last years, Porto has become an attractive tourism destination. According to Intercampus (2013b), the number of international visitors in the North of Portugal, where Porto is located, had risen up to 43% from 2004 to 2012. Tourism plays a crucial role in the Portuguese economy. According to the World Travel and Tourism Council (WTTC) in its 2012 economic report, the total contribution of travel and tourism accounted for 15.2% of the Portuguese GDP, representing 26.2 billion euros. These are striking results. It represent an opportunity to both public and private tourism players to maximize the benefits that tourism may bring to the region and to the country (Quilkin, 2013).

This study examines the image of Porto by analysing visitors' online publications and reviews in Travelpod - a travel blog website and in Tripadvisor - a travel reviews website. The aim of this project is to contribute to deeper the knowledge about Porto visitors' opinion. The analyses presented add value to agents and policy makers that work in the tourism sector in Porto by empowering them with a detail summary of what has been commented online and helping them assess their marketing strategies.

The paper is organized as follows. Section 2 and 3 present a literature review. Section 2 details the Destination Image concept, its formation and dimensions, while section 3 relates ICT to Tourism environment. Section 4 explains the research goal and the methodology used, including data collection and what kind of text mining tools supported the study. Section 5 presents the results. Section 6 concludes and suggests new research avenues.



## 2. Destination Image

*“The image and stability of destinations remains of critical importance.”  
in OECD (2014)*

The Destination Image theory is the main concept of this study. The definition of image is linked to the subjective perception of an objective reality that each one of us creates in our consciousness (Kesić & Pavlić, 2011). Using Reynolds (1965) words, page 69, “an image is actually the result of a complex process, it is the mental construct developed by the consumer on the basis of a few selected impressions among the flood of total impressions; it comes into being through a creative process in which these selected impressions are elaborated, embellished, and ordered”. Baloglu & McCleary (1999) note that since in the past decades studies have been suggesting that human behaviour is dependent upon image rather than objective reality, the interest and studies about image were developed in different areas of research such as tourism, marketing, business organisation and business strategy (Çakmak & Isaac, 2012).

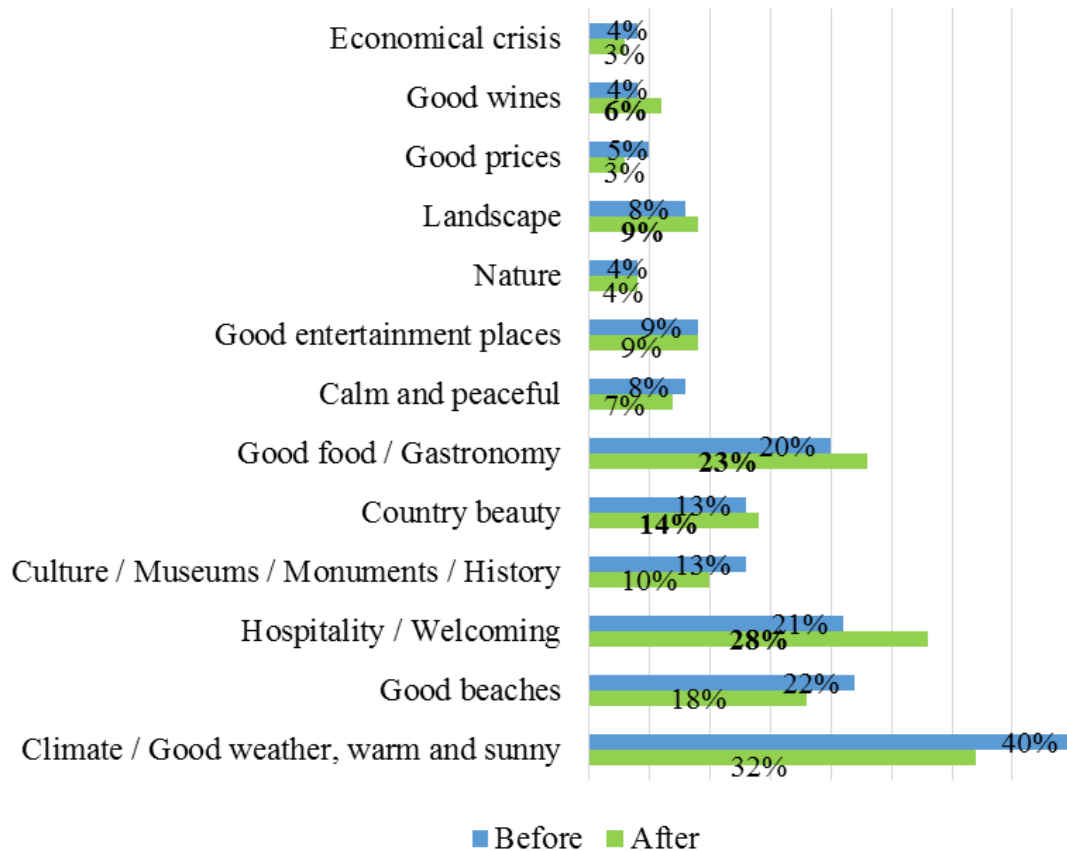
Destination Image concept emerged around 1975, when Hunt (1975), cit. in Hosany, Ekinci, & Uysal (2006), developed the notion of image in the tourism context. Destination Image reflects all objective knowledge and also impressions, preconceptions, imaginations and emotions that an individual might have of a particular place (Echtner & Ritchie, 1993). Those impressions can be correct or false, real or imagined, but they influence and shape a person’s behaviour (Barich & Kotler, 1991). Images change slowly and so it is crucial to regularly monitor travellers’ images (Gunn, 1988, cit. in O’Leary, 2005). Cultural distance also influences Destination Image. Consumers might be more confident in destinations with a culture similar to their own values (San Martín & Rodríguez del Bosque, 2008).

Country images are likely to influence people’s decision on purchasing, investing, changing residence and traveling (Kotler & Gertner, 2012). Destination’s image affects the behaviour of tourists in many ways, primarily as a key factor in the decision-making process of where to travel (Kesić & Pavlić, 2011). Secondly it influences the post-

purchase behaviour as it impacts the perceived quality, satisfaction, intention to return and willingness to recommend (O'Leary, 2005).

The increasing offer when choosing a tourism destination emphasizes the crucial importance of defining a unique identity and creating a competitive advantage. Motivations that resulted from a mixture of needs and desires, affect the propensity to travel. However, they do not influence the decision on choosing the destination (O'Leary, 2005). Images form the basis of the evaluation or selection process and thus provide the link between motivations and destination selection (Goodall, 1992, cit. in O'Leary 2005). Destination Images can be used as mental short-cuts for evaluating the information in the decision making process (Kotler & Gertner, 2012). Hunt (1975), cit. in Andreu, Bigné, & Cooper (2001), explains that images, beliefs and perceptions that individuals have about a destination may impact the growth of a tourist area as much, or even more, than tangible resources. Goodal (1992), cit. in O'Leary (2005), refers that when prices are comparable, image is the decisive factor in holiday choice. As a result, an image should generate interest and maintain a person's attention in order to convince the tourist to visit. The image should be simple, unforgettable, significant, sincere and relatively stable (Kesić & Pavlić, 2011). Each destination should carefully manage and treat its name as a brand (Kotler & Gertner, 2012).

No studies were found about Porto destination's image but there are some works published about Portugal. Three studies have investigated the online image of Portugal, (Walzem (2011) from United States, Quilkin (2013) from United Kingdom and Andrade (2013) from France). Also a survey study conducted by Intercampus (2013a) in Portugal discloses the words that better describe Portugal, before and after the trip experience (from a list of possible answers). The graphic 2 shows the main results. The most voted word/expression was climate/ good weather. It is interesting also to compare the evolution of results before and after the experience. Good climate, good beaches and also culture/ museums/ monuments/ history were less mentioned after the experience. While hospitality welcoming, country beauty, good food/ gastronomy, landscape and good wines are the answers that have improved its importance when describing the country after the experience.



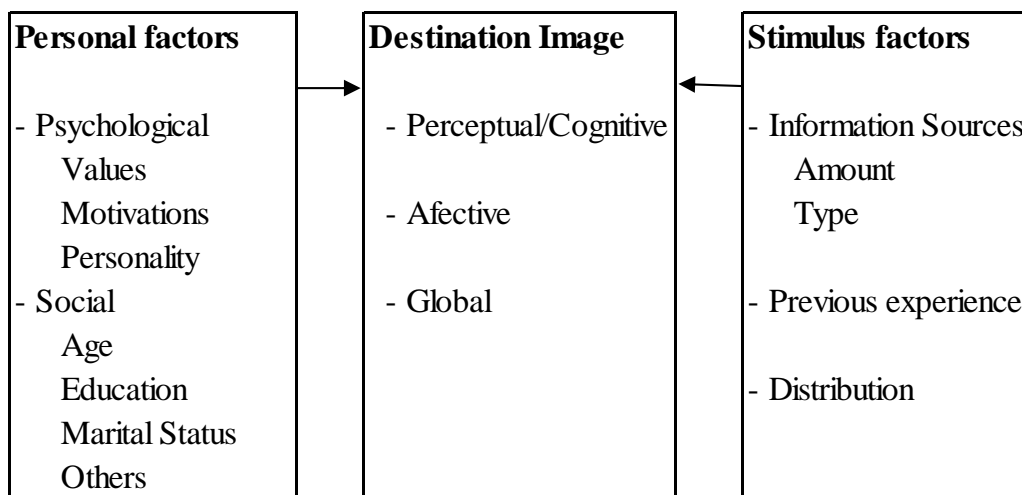
Graphic 2 - Word best describe Portugal  
Source: Adapted from Intercampus (2013)

After this introduction to Destination Image concept in the next part is exposed a general overview of the factors that constitutes the concept, how may be characterized its formation process and finally is exposed some possibilities to measure the concept.

## 2.1. Destination Image formation and dimensions

Destination Image formation is dependent on two kind of factors: stimulus factors and personal characteristics (Baloglu & McCleary, 1999). Image is formed from external stimulus and physical objects as well previous experiences. Personal factors, on the other hand, are the social and psychological characteristics of the perceiver, figure 1.

Figure 1 - A general framework of Destination Image formation.

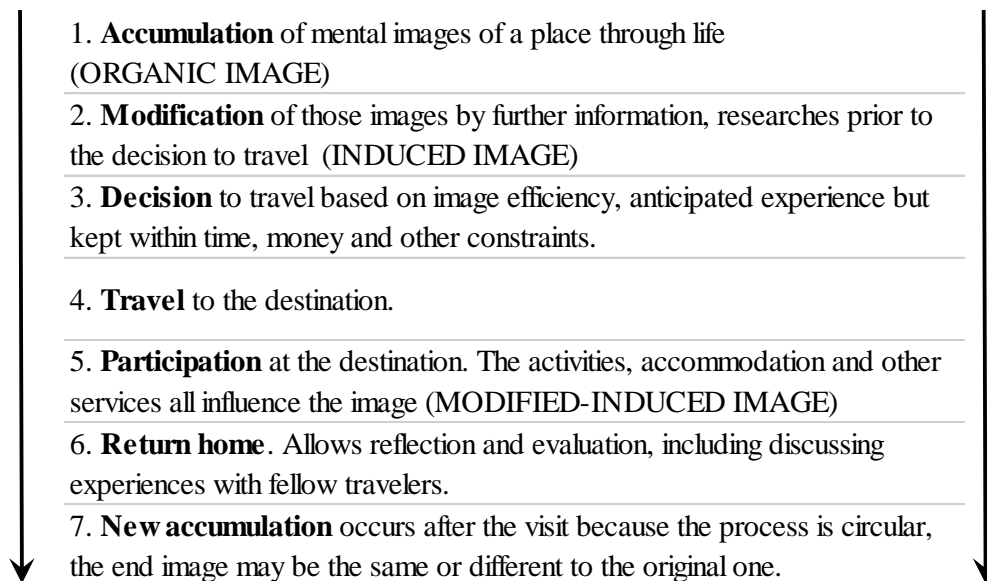


Source: Adapted from (Baloglu & McCleary, 1999)

A significant number of researchers also support that Destination Image is a multidimensional construction comprising of two primary dimensions: cognitive and affective (San Martín & Rodríguez del Bosque, 2008). The cognitive component refers to beliefs and knowledge about the physical attributes of a destination. While affective component refers to the appraisal of the affective quality of feelings towards the attributes and surrounding environment (Hosany et al., 2006). A study conducted by Kim & Yoon (2003) suggests that affective image components have more impact on developing Destination Image than cognitive components, especially feelings related to travel such as exotic atmospheres or relaxation. Nevertheless, cognitive components such as good restaurants, suitable hotels, friendly people and unique architecture are important when considering Destination Image attributes. Since tourists use both cognitive and affective dimensions to form their images on a destination, promoters should emphasize in the destination's positioning the physical and the emotions that is able to evoke in the tourist's mind (San Martín & Rodríguez del Bosque, 2008).

Gunn (1972), cit. in O'Leary (2005), was one of the first to identify the different ways in which cognitive images are formed through what is called Gunn's 7 - stages theory. This theory comprises a constant building and modification of images that are conceived at a number of levels namely organic, induced and modified-induced. In this 7 stages model, different influences and sources of information are organized in a sequential model of the travel experience (Echtner & Ritchie, 1993). The schema is presented in figure 2.

Figure 2 - 7 Stages model theory of Destination Image formation



Source: Adapted from Gunn (1972), cit in Jenkins (1999)

Echtner & Ritchie (1993) support and explain this path. There are three different stages of Destination Image formation. In phase 1, the image is called organic. It is based primarily upon information assimilated from supposedly unbiased sources such as books, documentaries or experiences of friends and family. In phase 2, promotional sources of information, such as travel agents or travel guidebooks, are used. The organic image changes to an induced image. The key difference is the control that the destination area has over what is presented; although little can be done to influence organic images, marketers can induce an image by investing in promotion (O'Leary, 2005). Finally, in the last stage, after consumer's own experience, image destination evolves for what is called modified-induced image. Doing an analogy of this model published in 1972 to the present, we may say that nowadays the stage 5 and stage 6 may happen at same time. With a easier

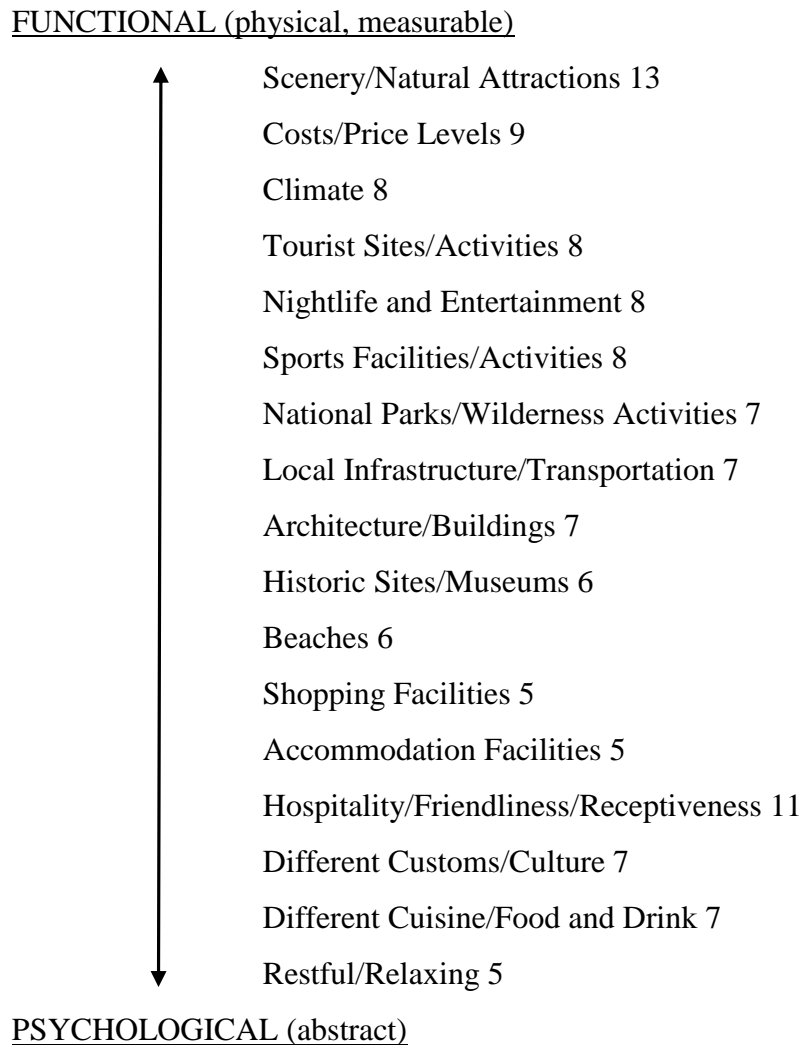
access to internet, the visitor does not have to return home to evaluate and share. Instead he/she may simultaneously experience the moment, evaluate and share with others.

There are also other aspects important to consider in Destination Image formation. Features can be defined in terms of both an attribute based and a holistic components (Echtner & Ritchie, 1993). Imagery is a way of process and store multisensory information based in holistic method of representing information, often described as mental picturing. On the opposite, attribute based is characterised by pieces of information on individual features or attributes (Echtner & Ritchie, 1993).

Some Destination Images can be adapted from directly observable or measurable characteristics, while others may be based on more abstract, intangible features. Therefore, the notion of functional and psychological characteristics, as suggested by Martineau (1958) in the context of retail stores, may be applied to Destination Image context. In this study, the author makes a distinction between the functional and psychological components of image when considering an entity such as a retail store. Functional characteristics are defined as directly observable or measurable (e.g. prices or attractions) whereas psychological characteristics cannot be directly measured (e.g. friendliness or atmosphere). Additionally, the author notes that both components are essential in determining the image of a store. It is possible to do a parallelism to the tourism context, being, for instance, the functional characteristic the cost of transports, restaurants and accommodations and the psychological characteristics the hospitality, safety ambience or local culture (Echtner & Ritchie, 1993).

It is not easy to define a list of attributes when designing a Destination Image. Echtner & Ritchie (1993) collect, and categorize, a list of attributes more used by other researchers, in a sample of 14 studies. Some of them are clearly functional, e.g. price, while others are psychological, e.g. friendship, and some lie in the middle. Figure 3 summarizes the attributes more used to measure Destination Image (the number in front of each attribute corresponds to the number of studies that had used it).

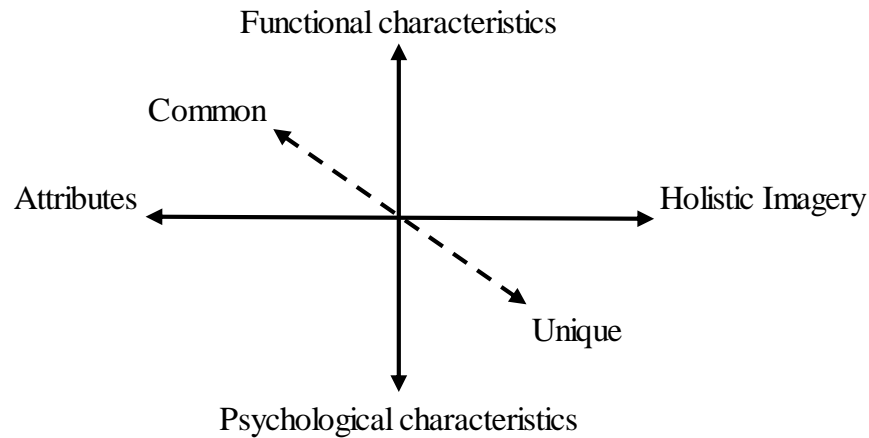
Figure 3 – Attributes to measure Destination Image.



Source: Adapted from Echtner & Ritchie (1993)

Echtner & Ritchie (1993) add also a new feature important to consider in Destination Image measurement. Concepts of common and unique (figure 4). While “common” components can be rated and compared between all destinations, e.g. price levels and infra-structures (as functional) or friendliness and quality of service (as psychological). On the opposite, “unique” is related to unique features or events of a destination, it also can be functional (e.g. India evokes the Taj Mahal), or psychological (e.g. Vatican associated to a set of values). Additionally, some destinations may also be identified by its atmosphere, e.g. Paris may be perceived as romantic destination.

Figure 4 - Destination Image components



Source: Adapted from Echtner & Ritchie (1993)

Figure 4 illustrates a resume of the features important to consider while measuring Destination Image (Echtner & Ritchie, 1993).

Consumers are not only perceivers of Destination Image information but they actively build and share their own images in the internet (Dwivedi, 2009). Because of the multiple dimensions of Destination Image (Gallarza, Saura, & Garcia, 2002) and the greater complexity the Internet has provided to destination marketing, investigating the Destination Image on the web is currently receiving great attention from researchers and destination marketing practitioners (Choi, Lehto, & Morrison, 2007).



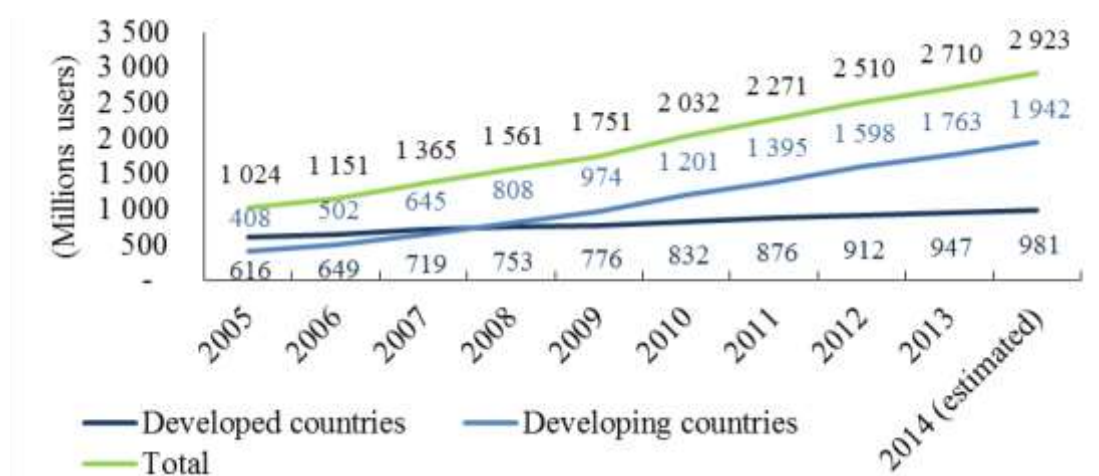
### 3. Travel 2.0.

*“Visitors are demanding more individual and authentic experiences and are relying more heavily on technology to plan and enhance their trips” in OECD (2014)*

This sector begins by explaining the impact of the ICT. Then, it explores the concepts of Web 2.0, social media and social networks and the role they play in changing the tourism sector, namely through a more active role of tourists in forming destinations’ images.

#### - ICT and the Tourism sector

According to International Telecommunication Union 2014, it is estimated that at the end of 2014, the number of internet users will almost reach three billion persons, which represents 40% of world’s population. Two-thirds of the world’s internet users come from developing countries (graphic 3).



Graphic 3 - Number of individuals using internet, 2005-2014 (estimated)

Source: Adapted from ITU World Telecommunication / ICT Indicators database

Tourism sector is also growing up. According to the World Bank statistics, the International inbound tourists (overnight visitors) have grown 50% in 10 years, from 718 million in 2003 to 1.078 in 2012. The revolution of ICT’s represented huge implications for the tourism industry (Buhalis & Jun, 2011). Travel planning is one of the main uses of the internet. A study conducted in Britain by Dutton & Helsper (2007) found that

respondents most frequent search activity online was making travel plans (84%), getting information about local events (77%) and looking for news (69%). Additionally, internet is establishing itself as an important place to buy travel services, representing already 54% of bookings, while travel agents represent 24% (OECD, 2014). Traditional operators are facing a new consumer, which can easily access information and share photos, comments or suggestions in an informal and collaborative way, influencing other consumer decision making (Miguéns, Baggio, & Costa, 2008). Worldwide larger access to internet has shifted the power from the providers of tourism services to consumers, and eased a change to self-guided holidays and independent travel (OECD, 2014). Consumers are no longer dependent on web site owners to find the information they seek, as they can increasingly rely on unfiltered, dynamic and topical information provided by their own peers. Instead of the marketer dictating how information is presented and consumed, the user is now in control.

#### **- Web 2.0**

The concept Web 2.0, developed by O'Reilly (2005), refers to “the second-generation of Internet-based services that let people collaborate and share information online in perceived new ways—such as social networking sites, blogs, wikis and communication tools” (Lai & Turban, 2008). In this era, users add tremendous value (O'Reilly, 2005). The website is no longer a static page, but a dynamic place which allows users to generate content with an easiness interactivity promoting the formation of communities and the generation of user-driven content (Miguéns et al., 2008). Web 2.0 applications in the tourism sector have been named Travel 2.0 and are creating a cultural change in the tourism world (Miguéns et al., 2008). Consumers not only take an active role in choosing and preparing their own trips, but also helping other tourists in forming an image or idea of their destination before travelling (Muñoz, Hernández, & Sánchez, 2012).

#### **- Social Media**

This new reality takes us to the concept of Social media, defined as the online platforms and tools used to share opinions and experiences through text, photos, videos, music, insights, and perceptions with each other (Lai & Turban, 2008). Social media platforms, such as Twitter, Facebook, YouTube along with the influence of sites such as Tripadvisor,

have moved the source information to a powerful online community often with local knowledge of tourism products and services and within an immediate feedback (OECD, 2014). As an influential democratic force, social media enables people, rather than organizations, to control and use various media easily at little or no cost; consequently, it enables communication and collaboration on a massive scale (Lai & Turban, 2008).

#### **- Social Networks**

Then there is also the concept of Social Networks, where web tools allow individuals to construct a public or semi-public profile, share a connection with other users, view and navigate their list of connections and those made by others within the system (Boyd & Ellison, 2007). People may create their own space or home page (Lai & Turban, 2008), connect in social networks and share common interests and activities. Social networks provide to users a vast interaction options, such as chat, video conferences, participation in blogs or discussion groups (Miguéns et al., 2008). According to a survey conducted at the World Travel Market Trade Fair held in London in 2010, 36% of tourists from the UK used social networks as a source of information for making travel arrangements (Muñoz et al., 2012).

Transparency, community, personalization and experience are the buzzwords associated to Travel 2.0. Official websites no longer satisfy the current needs or interests of consumers in terms of planning their trip (Conrady & Buck, 2007).

### 3.1. Electronic Word Of Mouth (e-WOM)

Word Of Mouth (WOM) is defined as an informal communication occurring in person, by telephone, email, mailing list, or any other communication method about a service or a good (Goyette, Ricard, Bergeron, & Marticotte, 2010). A recommendation source may be personal or impersonal (Goyette et al., 2010) and can be established between producers and consumers as well as between consumers themselves (Litvin et al., 2008).

More recently, within Web 2.0 era, WOM has developed to electronic-Word of Mouth (e-WOM) which is equally defined as communications directed at consumers but through Internet-based technologies (Litvin et al., 2008). E-WOM communication possesses unprecedented speed of diffusion and enables multi-directional exchanges of information between communicators and receivers (Goyette et al., 2010). Consequently, e-WOM is revolutionizing consumer's behaviour (MacKinnon, 2012) and transforming travel industry. Studies in the tourism area demonstrate the influence of both positive and negative WOM upon tourism products and additionally confirm its importance in increasing travel intentions, creating a favourable image of the destination, and reducing promotional expenditures (Jalilvand & Samiei, 2012).

### 3.2. User Generated Content (UGC)

The popularity of Web 2.0 has empowered consumers to influence others through a User Generated Content (UGC) tools such as blogs, forums or social networking sites (Cheung & Thadani, 2012). UGC and e-WOM are two important concepts and often confused. The two differ depending on whether the content is generated by users or if it is transmitted by them. E-WOM might be originated by a power position, for example a marketer; while UGC is created from the consumers themselves, people who voluntarily contribute with information. For example, if a tourist posts on [www.youtube.com](http://www.youtube.com) a video about his holidays in Porto, that is UGC. However, when internet users share the video with their friends it is also e-WOM. Though UGC and e-WOM are distinct concepts but closely

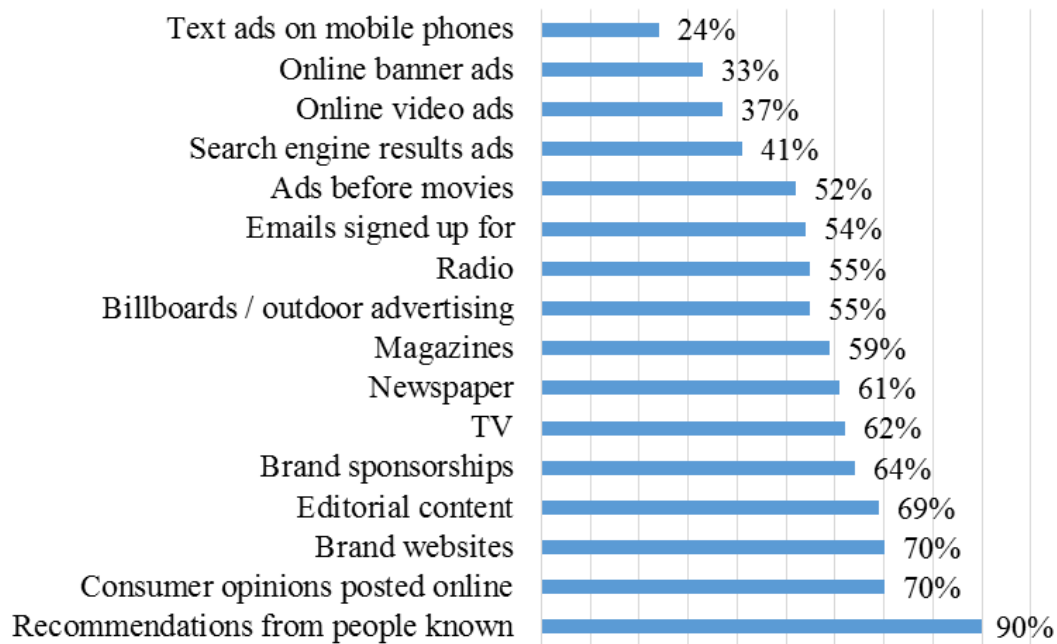
related. To be successful, e-WOM depends on the dissemination of content, and UGC has less influence without e-WOM (Cheong & Morrison, 2008).

The potential impact of consumer reviews can be significant, due to the specific nature of holidays. Holidays are intangible products, difficult to evaluate before acquisition, characterized as high involvement and risk considering the personal importance usually associates (Papathanassis & Knolle, 2011).

The growth of User Generated Content is clearly affecting travel consumer decisions. Most readers perceive travel reviews are being more likely to provide up-to-date, enjoyable and reliable information in comparison to what is provided by travel service providers (Ulrike Gretzel & Yoo, 2008). Consumer reviews serve two distinct roles: provide information about products and services and serve as recommendations (Do-Hyung, Lee, & Han, 2007). Reviews are particularly important for the accommodation product, with relevance for other travel products much smaller (Ulrike Gretzel & Yoo, 2008). Trip characteristics seem to play a key role when it comes to travellers' engagement with UGC for travel planning (Zeng & Gerritsen, 2014), Simms (2012) found that a higher percentage of travellers turned to user-generated content when visiting a destination for the first time, as well as when visiting an international destination (Zeng & Gerritsen, 2014).

UGC sites can be divided into several categories such as: social networks (e.g. facebook, linkedIn), blogging (e.g. travelblog, twitter), forums (e.g. travelforum, lonelyplanet), video-sharing (e.g. youtube), photo-sharing (e.g. flickr) or review-sites (e.g. Tripadvisor) (Davydova, 2012).

A study published by Nielsen (2009) – graphic 4 - concluded that after recommendations from people that are known, the most trustful sources were online consumer's opinion follows with 70% percent of respondents trusted “completely” or “somewhat” recommendations from consumer opinions posted online.



Graphic 4 - Degrees of trust per type of advertising  
Source: Adapted from Nielsen Global Online Consumer Survey (2009)

A survey conducted by Ulrike Gretzel (2007) sponsored by Tripadvisor, with a sample of 1480 persons, concluded that most of the surveyed Tripadvisor.com users do not like to make spontaneous decisions and plan their trips to a large extent in advance. Almost all participants use Internet as an information source for planning pleasure trips and a majority (82.5%) uses it every time they plan a pleasure trip. Their most frequent travel planning-related online activities are: looking at other consumers' comments/materials (90%), printing out maps/directions (82.7%) and reading travel-related blogs (64.2%). More than half (57.8%) read other travellers' online reviews every time they plan a pleasure trip while 36.7% read them very often or frequently. More than three quarters (77.9%) of online travel review readers think that other travellers' reviews are extremely or very important for determining where to stay. Additionally over 80% confirm that reading other traveller's online reviews increases confidence in decisions, makes it easier to picture what a place would be like, helps reduce risk and uncertainty, makes it easier to reach decisions, and helps with planning pleasure trips more efficiently.

One other study (Intercampus, 2013a) developed through questionnaires, considering a sample of tourists in Portugal also concluded important numbers: about 36% of inquiries chose Portugal for tourism destination from family, colleagues or friends recommendations; and 22% from information seen in internet (this value varies from country of origin, Spain is the nationality where information seen in internet has a bigger weight: 34%).

The Web is evolving from a business-to-consumer marketing media to one where peer-to-peer generation and sharing of data are the norm (Peter O'Connor, Höpken, & Gretzel, 2008). Though, the tourism industry is required to treat consumers as co-producers and leverage network resources to successfully operate their businesses in this consumer-centric era (Buhalis & Jun, 2011). Tourism businesses are facing the need to implement strategies and tools (websites or portals) based on User Generated Contents or, at least, to incorporate these new technologies to enrich their multimedia contents (Miguéns et al., 2008). Organizations should utilize online reviews to improve their understanding of the market's reaction to their offering and subsequently feed this information into their product development and quality control processes (Dellarocas, 2003). Through positive online feedback new customers can be acquired and retained at a relatively low cost, while negative feedback can obstruct customer acquisition and reduce the value of brand in the longer term (Dellarocas, 2003).

### 3.2.1. Weblogs

One of the most highly touted features of the Web 2.0 era is the rise of blogging. Personal home pages exists since the early days of the web (O'Reilly, 2005). Internet has become an important medium for sharing visitor's storytelling (Martin, Woodside, & Dehuang, 2007). First-hand visitor reports experiencing destinations indicate that tourists tell stories that offer clues of how they interpret and enact the myths that these destinations enable (Martin et al., 2007). Tourists are looking after experiences and that is one of the aspects

of tourism: its experiential element; that fact aligned with the desire on the part of some tourists to share theirs, has created the perfect conditions for the growth of tools such as travel blogs (Prats & Marín, 2012). Stories do move people profoundly (McKee, 1997), they have the potential of influencing a substantial number of future visitors who go online in search of first-person unbiased reports of others' own destination experiences (Martin et al., 2007). Travel blogs are becoming an increasingly important mechanism for exchanging information among tourists, and for marketers to learn about the attitudes of their markets (Wenger, 2008).

Following Thevenot (2007) research, blogs can be classified into three main categories: individual, collaborative and corporate. In Individual blogs (or personal), the writer's combine personal life and a certain subject that they would like to provide information or comments on. Some of the most well-known nowadays include [travelblog.org](http://travelblog.org) and [travelpod.com](http://travelpod.com), [mytripjournal.com](http://mytripjournal.com). There is also a space for personal travel blogs on specialized customer review sites (e.g. [Tripadvisor.com](http://Tripadvisor.com)) and other travel related companies' sites (such as travel agencies and travel guides). In collaborative blogs, posts are written by more than one author. This makes those blogs more regularly updated, having wider range of information and whereby they are seen as being more relevant for the reader. While in personal blogs the author can choose any topic he wants for the next post, collaborative blogs has a specific topic which has to be followed. Finally, corporate blogs are the ones published by a certain organization.

This research will study reviews posted in [www.travelpod.com](http://www.travelpod.com), an individual blog. In this webpage the blogger creates his own blog reporting different topics he desire. The reader has the possibility to search posts from one blogger or search for a specific topic from all writers. That was how data was collected, searching posts related to Porto.

More and more studies have found that consumers tend to be more interested in products discussed on blogs and web forums than they are in those marketed in traditional ways. E-WOM can thus contribute raise the credibility of a product or service (Bickart & Schindler, 2001). It remains a challenge for managers to incorporate e-WOM effects into their business planning (Rong, Vu, Law, & Li, 2012), it is fundamental to focus in this type of communication to obtain a significant competitive advantage. In times where consumers trust on organizations and advertising reduced and a decrease in television



advertising is visible (Jalilvand & Samiei, 2012), key stakeholders in tourism will need to develop the competencies to cope with emerging digital and social media shifts. The scale and pace of these developments creates challenges to continually update marketing strategies, including online strategies (OECD, 2014), the industry winners will be those quick enough to adapt their business plans to the new dynamics of the online marketplace (Conrady & Buck, 2007).

In the last two sections the fundamental theoretical concepts of this research are presented. In Destination Image, after an overall definition, the formation process and the main attributes defended by several authors were described. In Travel 2.0, the complex mutation in ICT's, emphasizing the concepts and impacts of UGC and e-WOM. The importance of those two concepts to the success of a tourism destination is undeniable. Tourism marketers should be focused on that. We expect this investigation shed some lights towards in that way.

## 4. Research Goal and Methodology

As explained, online recommendations have a big weight on what concerns image formation and purchase decision making process. Studies also show that nowadays user generator content is seen as more reliable than official sources (graphic 4, page 16). Thus, the goal of this study is to uncover online consumer-based image of Porto, e.g. the image formed by the aggregation of opinions expressed online by tourist after experiencing the city.

The reasons behind the choice of Porto for this study are related to the notable growth and development seen in the past years. According to national statistics (Intercampus, 2013a), in 2004 the North of Portugal region received 1.834.017 tourists and in 2012 that number grew up to 2.626.472 (an increase of 43%). There are several reasons that may help to understand why Porto and the North of Portugal are becoming more attractive and receiving more tourists. Although this is not the aim of the paper, it is important to contextualize and shed a light in what has been contributing to the success of Porto as a tourism destination. Ryanair, the low cost airline company, has been playing an important role in raising the number of visitors. Currently there are approximately 40 Ryanair routes from Porto to 8 different countries in Europe. The increasing number of flights has been coupled with an increasing number of infra-structures, as accommodation, restaurants and entertainment. In the past years Porto is also winning many prizes such as “Top 10 European travel destinations for 2013” by Lonely Planet - world’s largest travel guide book publisher (LonelyPlanet, 2013) or “Top 10 destination on the rise - Europe in Traveler’s Choice 2013” at Tripadvisor (PubliTuris, 2013). This study intends to find some insights of what may contributed for this apparently good phase Porto tourism is passing through. We hope to bring a new point of how tourists see Porto and how they describe their experience.

According to Jenkins (1999), researchers should consider to use a combination of structured and unstructured methods to measure multi-components; in the past, scholars tended to use only structured methods to measure Destination Images. By using structured methods, using e.g. Likert scale, a researcher asks tourists to rate a set of attributes and applies statistical methods to explain a group of tourists’ Destination Image. This kind of

study is being criticized as Destination Image structure is highly dependent upon the researcher's conceptualization and reflects the researcher's own perceptions and dimensions (Çakmak & Isaac, 2012). Attributes-based research is also said to categorise consumers excessively as rational thinkers rather than storytellers using the narrative mode (Govers and Go, 2003 cit. in Dwivedi, 2009).

In 2002, after reviewing 142 papers, Pike (2002) concluded that less than half of them used any qualitative methods involving customers' opinions. However, with the Web 2.0 revolution, the available text and visual data on the web increased exponentially and helped researchers to use qualitative assessments like content analysis and quantitative methodologies as word frequency studies (e.g. Choi et al., 2007, and Son, 2011). This study will focus on the most frequent expressions tourists write when posting in a blog or when reviewing an experience on a tourism website. In this study, methodology will be focus in a quantitative method, investigating the more frequent words used by reviewers when reporting their travel experience in Porto.

Digital data is now everywhere - in every sector, economy or organization (McKinsey, 2011), the need to understand large and complex information affects all areas. In the business world, corporate and customer data are becoming recognized as a strategic asset. The capacity to discover information hidden in these huge data and act on that knowledge is increasingly important nowadays (Kantardzic, 2011). Consequently, new forms to collect and analyse data should be considered. Text mining is one example. Text mining is the process of relevant and focal knowledge from text documents. Tan (1999), page 65, describe "text mining as a multidisciplinary field, involving information retrieval, text analysis, information extraction, clustering, categorization, visualization, database technology, machine learning, and data mining".

## 4.1. Data collection

In this research, data collection is based on a non-participatory method since researchers did not interact with the sample. Data was collected from two different kind of sources: blog website [www.travelpod.com](http://www.travelpod.com) and the popular travel reviews website [www.tripadvisor.com](http://www.tripadvisor.com). In blogs, literature is more like storytelling, the author has freedom to comment anything, on the other hand, in Tripadvisor it tends to be a shorter and more direct speech about a specific attraction or service but the intervenient is just able to comment topics/places that are categorized in the website. Two different sources were selected to benefit different kinds of information and, hopefully, allow a more complete illustration of online's Porto image.

Blogs sources were chosen through a search on the 8<sup>th</sup> April at [www.google.com](http://www.google.com) with the sentence "travel blog". The first option was excluded because it did not contain data about Porto, the second option was taken: [www.travelpod.com](http://www.travelpod.com). All reviews written in English available on 12<sup>th</sup> September were considered, totalizing 519 posts.

Tripadvisor was chosen not just because it appeared in first position when searching "travel guide" at [www.google.com](http://www.google.com), as advertisement, but also because it is the world's largest travel site (comScore, 2013, cit. in Tripadvisor). Tripadvisor offers advice from real travellers and a wide variety of travel choices and planning features. It offers also information about hotels, vacation rentals, flights, attractions, restaurants, forum's and plenty of photos (mostly from reviewers). It has nearly 280 million unique monthly visitors (IpsosMediaCT, 2013), 170 million reviews and opinions covering more than 4 million accommodations, restaurants and attractions (Tripadvisor, 2014). The website operates in 44 countries worldwide. For this analysis we focused on reviews related to Attractions, Accommodation (Hotels, B&B and Specialty Lodging) and Restaurants in Porto District. Data was collected also on 12<sup>th</sup> September 2014. Due to space limitation of the freeware software used in data analysis, a maximum of 350 reviews per establishment or place were collected in Attractions and Accommodations, totalizing 2924 and 4413 reviews, respectively. In Restaurants collection, due to a large number of restaurants, a maximum of 50 reviews per each were selected, totalizing 4748.

## 4.2. Data analysis

The methodology chosen in this study is a little different from the previous researches related to Destination Image of Portugal found. Data analysis focus on a quantitative method, through the frequency of words. Each analysis in the Results section starts with a tag cloud. Tag cloud is a weighted visual list, in which words that appear most frequently are larger and words that appear less frequently are smaller, allowing a fast perception of the most relevant concepts in a large body of text (McKinsey, 2011). After that illustration, results are analysed in more detail. For a more complete analysis, tag clouds are created in bigrams. Bigram is a pair of consecutive words. With this option we hope to find more completed conclusions. Nevertheless, a list of the top frequency words and bigrams are also presents in annexes. Some interpretations are made, sometimes searching the context involving the word directly in reviews text. After that, bigrams are categorized by nationality, presenting countries with more reviews in English and are made some comparisons. Additionally, some words or bigrams were excluded due to no relevance for the study; normally a word over repeated by one reviewer that might influence the overall results.

This study is developed with text mining software, the free version RapidMiner (Hofmann & Klinkenberg, 2013). It requires no programming, just build processes with a drag and drop interface. At the end, data is presented in a simple and easy visualization through tag clouds. Figure 5 describes, in a simple scheme, how the RapidMiner process works in four steps:

Each phase of figure 5 is briefly described in the following lines (detail image of each phase in annex 1).

Figure 5 - RapidMiner process resume



Source: RapidMiner software

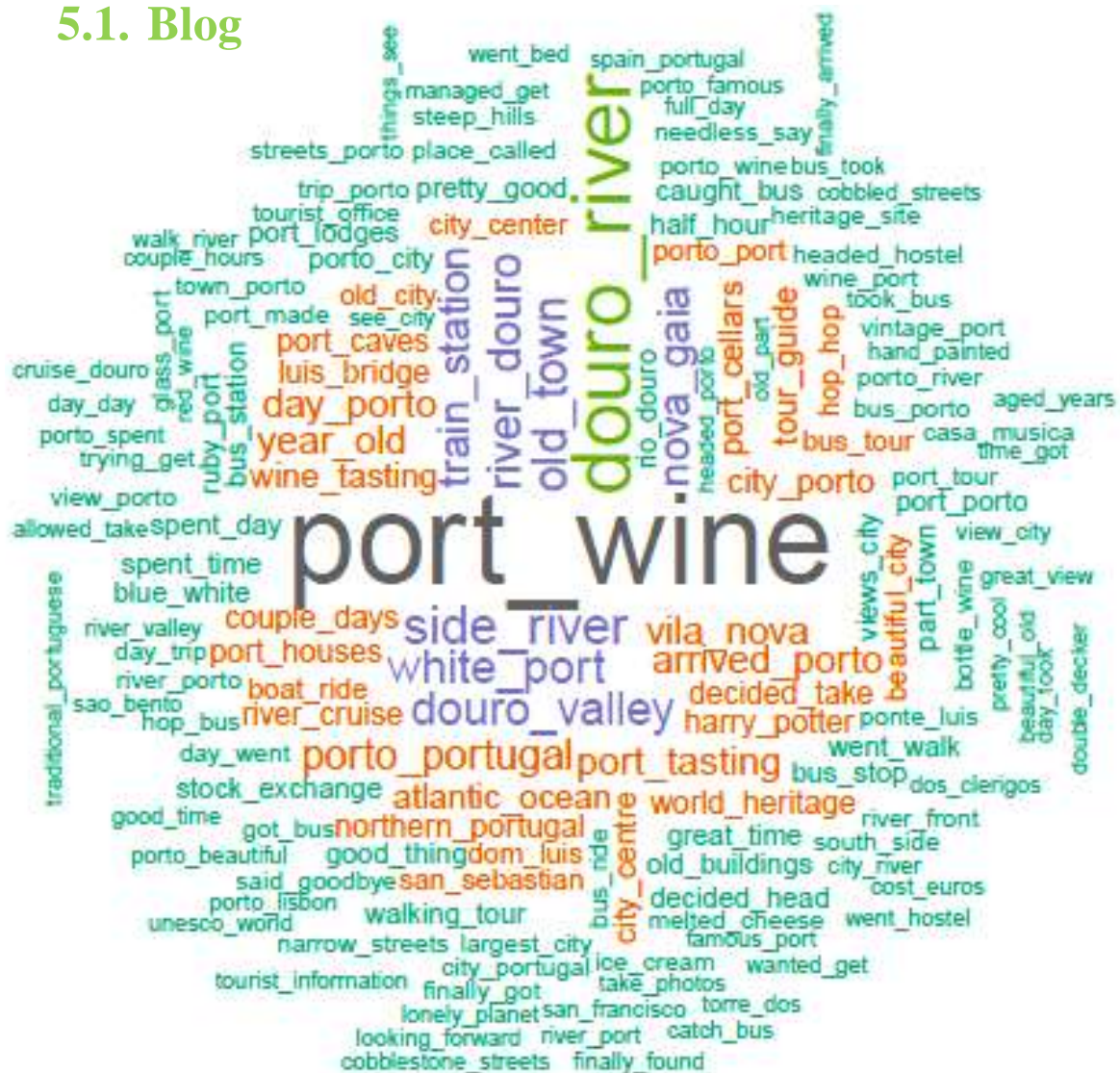
1. Crawl Web - In this step the link is included for the websites we want to collect the data from. In this study, there were necessary four different collections: [www.travelblog.com](http://www.travelblog.com) searching blogs posts about Porto and in [www.Tripadvisor.com](http://www.Tripadvisor.com), after selection Porto (Porto District, Northern Portugal, Portugal) three subsections were collected: attractions, restaurants and hotels.
2. Process Page - In this stage is where it is specified which data to collect and the way to find it. There was collected data like the title, nationality of the writer, the testimonial or review, and in the case of Tripadvisor the evaluation rates in what concerns restaurants and hotels.
3. Process Review - In this step, after defining the topics to collect, we specified what kind of data we want to pull. In this case, we excluded stop words rejecting words with less than 3 letters, searched by bigrams (pair of consecutive written words).
4. Word Cloud - In this final step, tag cloud characteristics such as size of words or colour are defined.

Let us pass to next section, to discover the main results.

## 5. Results

This part is organized in four sub-sections, each one referred to a different source of data: Blog and Tripadvisor divided by attractions, restaurants and accommodation (include hotels, bed&breakfast and special lodging). Each section starts with a tag cloud, presenting the more frequent bigrams, followed by a more detailed of each topic. Additionally an analysis by country of origin is made, some comparisons are study between top frequency bigrams of the three nations with more comments written in English.

## 5.1. Blog



### Illustration 1 - Blog cloud

A total of 519 reviews were considered in TravelPod. The size of the words in the cloud is closely related to the frequency of the word used by Travelpod writers. In the centre, in grey, it is the most used word. The subsequent levels of importance are in green, violet and orange. There are several messages that can be extracted from this tag cloud.

There is a widely recognition of associating Porto to PortWine. Although the cellars are not located in Porto itself, Port wine is attracting the interest of Travelpod writers. There are several bigrams in different colour layers that support the importance of Port-wine, such as port tasting, wine tasting, port cellars, port houses or port caves. White port is also highlighted, it is not well known as the red port and some visitors discover it in their trip. Also Nova Gaia referring to the city where the wine cellars are installed - Vila Nova de Gaia.

The Douro River also attracts some attention. Cruises and the boat rides are mentioned several times. Douro word is also sometimes associated to Douro wine and Douro valley.

There is also a common notion of seeing Porto as an old town. Many references are made to degradation state of some buildings. However, its monuments are not widely cited. Luis Bridge, and the train station are amongst the most highlighted. D. Luis I bridge (named as Luis Bridge) is very appreciated by its iron architecture drawn by Gustavo Eiffel. The train station is often related to S. Bento train station with its appreciated blue and white tiles works. Harry Potter is also much cited referring to a touristic bookstore where is said to inspire the Harry Potter author to write the book. The Stock Exchange Palace (Palácio da Bolsa) follows in top frequencies. On the contrary, Torre dos Clérigos, the Church of S. Francisco or Casa da Música, seem to have been overlooked despites being usually suggested in tourist guides. There are also few remarks on the Portuguese cuisine. Tour guide is also highlighted. Many visitors refer to use a tour guide to visit the city and also in the wine cellars visit.

In table 1, a synthesis top 10 bigrams by country of origin (more data in annex 2). Those are the countries with more posts published in English. It is unquestionable that the main bigrams bloggers use to describe their experience in Porto are related to the Port wine, the river and title Porto as an old town. Those words appear in different contexts with a strong incidence across all nationalities. A particularity in Americans' posts is the



expression San Sebastian. Further searching of these reviews revealed that many visitors include San Sebastian (a city located in the North of Spain) in their trip itinerary.

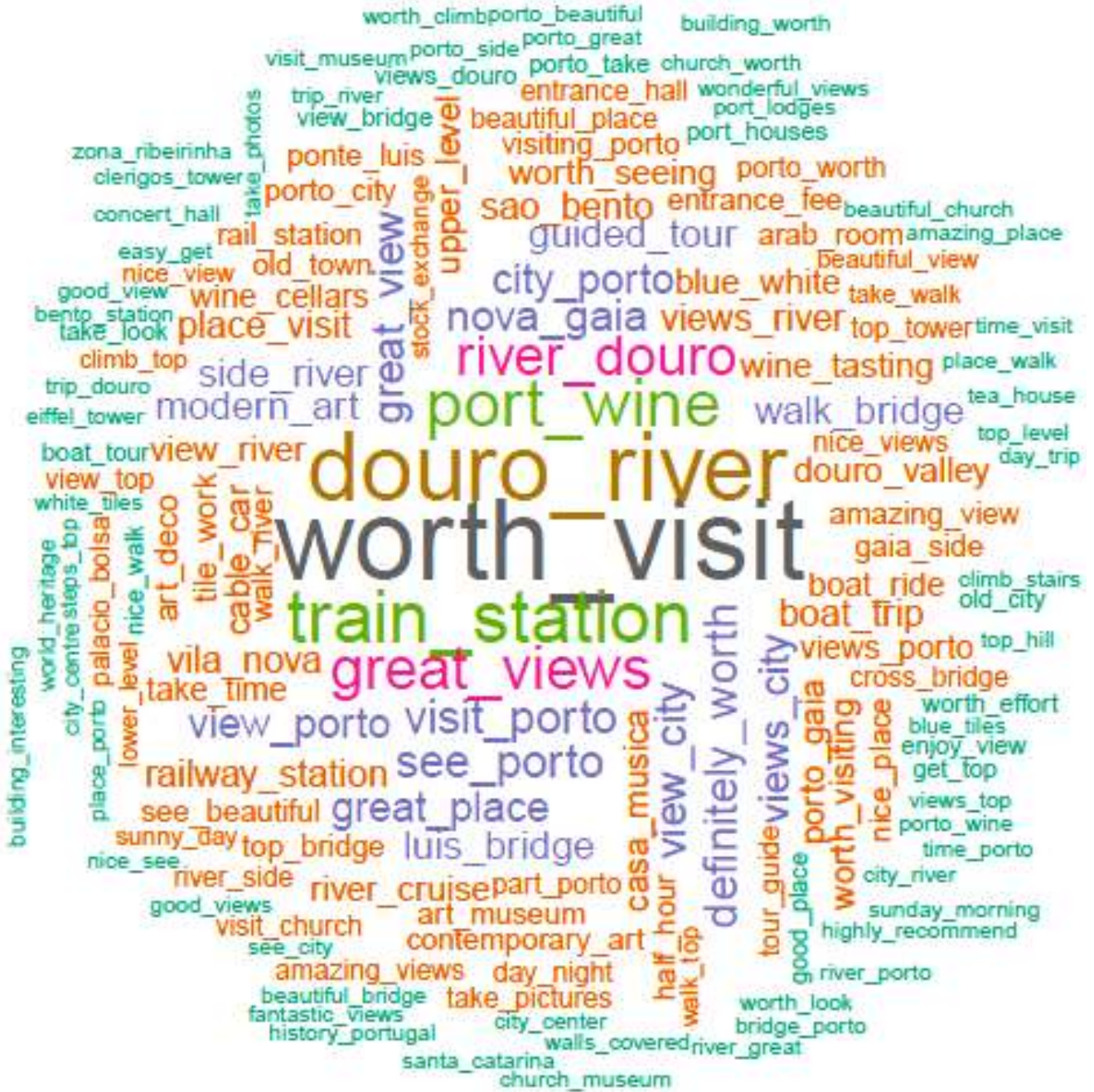
	<b>United States</b>	<b>Freq.</b>	<b>Australia</b>	<b>Freq.</b>	<b>Canada</b>	<b>Freq.</b>
1	port_wine	52	port_wine	31	port_wine	25
2	douro_river	26	old_town	20	douro_river	22
3	arrived_porto	13	train_station	19	side_river	14
4	porto_portugal	13	douro_river	18	douro_valley	11
5	san_sebastian	12	said_goodbye	14	old_town	11
6	city_porto	11	port_tasting	13	train_station	11
7	side_river	10	white_port	12	year_old	11
8	boat_ride	9	bus_station	11	day_porto	10
9	douro_valley	9	day_porto	11	white_port	10
10	pretty_good	9	headed_hostel	11	nova_gaia	9

Table 1 – TravelPod - top bigrams by nationality

After this analysis of the main results in Blogs testimonials we will now start the analysis in Triadvisor reviews.

## 5.2. Tripadvisor

## Attractions



### Illustration 2 - Attractions cloud

A total of 2924 reviews were considered in this subsection. Again, the colour and the size of the words refer to their frequency on those reviews. This cloud highlights the attractiveness of Porto. By outlining its worthiness, Porto is formally recognized as a tourism destination. Important is now to understand what are the most valuable sights of the city. Interestingly is to note that the words used by Tripadvisor are in line to those used by Traverpod writers. Porto wine and Douro river are again taking the central role. Also in terms of monuments, D. Luis I bridge and S. Bento train station are widely mentioned. The views are also highlighted, when searching in visitor's posts it is possible to see its association with the river and is frequently mention the views from and of the bridge. In addition, there is a reference to other important monuments or museums as Casa da Música, the Stock Exchange Palace (or Palácio da Bolsa) and references to modern and contemporary art concern Serralves museum, which deserves here a prominent mention. The river seems privileged over the sea, the sea and the beaches are not acknowledged as Porto's main attractions once there is no mentions in top 150 words or bigrams. In order to gain a deeper understanding of the image Tripadvisor writers have from Porto, an analysis to the more frequent words was also analyse. Adjectives with higher frequency have a positive connotation, such as beautiful, great and nice. These are outstanding results to the city of Porto and corroborate the analysis of the tag cloud, where the most important words were "worth\_visit".

Table 2 presents the top ten more frequent set of bigrams segmenting them by nationality. The majority of the English contributions on Tripadvisor Attractions come from the United Kingdom, United States and Canada. All three highlight its worth visit and Douro river. Looking for some differences, in a deeper list of 30 bigrams available in annex 2, we can identify that United States visitors highlighted some monuments that other two did not. Casa da Musica, Stock Exchange and Art Museum are only referred by United States. On the other side, United Kingdom is only referring Cable Car and Canadians are the only refereeing Clerigos Tower.

	<b>United Kingdom</b>	<b>Freq.</b>	<b>United States</b>	<b>Freq.</b>	<b>Canada</b>	<b>Freq.</b>
1	worth_visit	47	douro_river*	34	see_porto	5
2	river_douro*	29	train_station	31	take_time	5
3	great_views	26	great_views	18	worth_visit	5
4	side_river	15	worth_visit	16	tile_work	4
5	views_river	15	river_cruise	15	train_station	4
6	railway_station	14	boat_ride	12	view_city	4
7	definitely_worth	12	walk_bridge	12	clerigos_tower	3
8	port_wine	12	definitely_worth	11	douro_river	3
9	views_city	12	port_wine	11	douro_valley	3
10	worth_visiting	12	casa_musica	9	gaia_side	3

Table 2- Tripadvisor Attractions - top bigrams by nationality

(\*) The bigrams River\_Douro and Douro\_River were aggregated once just the words order is changed, keeping the same meaning.

Despite the lack of reference to the Portuguese cuisine and the accommodation services on the previous contributions, Tripadvisor writers can make separate contributions on these two topics. The paper will analyse what Tripadvisor writers have been writing regarding the restaurants and hotels in Porto.



## Restaurants



Illustration 3 - Restaurants cloud

A sample of 4748 reviews was considered in the analysis of Tripadvisor restaurants. Once again, there is an overall positive idea about the restaurants in Porto in terms of quality of service and its cuisine.

Tripadvisor contributors noted the quality of the food and compliments as delicious, tasty and fresh are observed (anexx 4). When searching words related to drinks and food, wine is also here the most cited word. Fish is mentioned more times than meat. Surprisingly sea bass have more frequency than cod fish, as the last one is seen by locals as one of the more traditional elements of Portuguese gastronomy (anexx 4). Many highlights are seen describing the cuisine as Portuguese and typical. Porto's famous dish, Francesinha and olive oil are also mentioned. Sushi is an exception to the traditional trend seen, as it is not part of Mediterranean diet (in which Portugal is included).

Service and price-quality ratio are also highly commented. Positive descriptions as friendly staff and good service are generally comment. When mentioning costs, the highlighted bigrams suggest also positive feedback considering reasonably price.

Table 4 compiles the frequency of the words by country of issuance/publication. The countries with more posts in English were United Kingdom, United States and Canada. The table clearly outlines the quality of the Portuguese food. It is interesting to realize each that United States reinforce in the same level the food and service, while the other two highlighted first the gastronomy.

	<b>United Kingdom</b>	<b>Freq.</b>	<b>United States</b>	<b>Freq.</b>	<b>Canada</b>	<b>Freq.</b>
1	food_good*	78	service_good*	35	food_good*	17
2	good_value	54	food_good*	35	food_excellent	9
3	service_good*	40	highly_recommend	20	service_excellent	8
4	staff_friendly	31	food_excellent	13	chocolate_cake	5
5	wine_list	24	staff_friendly	13	food_delicious	5
6	main_course	23	house_wine	12	glass_wine	5
7	quality_food	23	food_great	11	service_good*	9
8	reasonably_priced	22	trip_advisor	11	wine_list	5
9	night_porto	21	friendly_helpful	10	great_service	4
10	traditional_portuguese	21	great_food	10	reasonably_priced	4

Table 3 - TripAdvisor Restaurants - top bigrams by nationality



## Accommodation



### Illustration 4 - Accommodation cloud

In “Acommodation” category, 4413 reviews were considered from Tripadvisor’s hotels, bed&breakfast and special lodging sections. Once again, a highlighted point is the friendly and helpful staff. One example of that is the impressive frequency the bigram “Carmen joao” have. Those are the two hosts of a guesthouse located in the riverside, many of their guests are fascinated with their service.

The location is also frequently commented. In general guests seem satisfied with the accommodation location, words as walking distance, metro and train station also appear with a high presence. Some other compliments such as clean rooms and good breakfast appear in top frequencies. Related to price, tourist seem also satisfied and consider a good relation value money.

Analysing top 10 bigrams by nationality in the Table 5, we may visualise that results are similar, all three nationalities highlights location, the friendliness and help provided by their hosts, and also cleanliness and breakfast. When searching some more information in anexx 5, it is seen that United Kingdom is the one most referring features as air conditioning and wi-fi.

	<b>United States</b>	<b>Freq.</b>	<b>United Kingdom</b>	<b>Freq.</b>	<b>Canada</b>	<b>Freq.</b>
1	guest_house	84	stayed_nights	56	guest_house	37
2	carmen_joao	56	guest_house	48	house_douro	20
3	train_station	38	place_stay	47	walking_distance	15
4	house_douro	35	staff_friendly	47	carmen_joao	11
5	place_stay	33	metro_station	41	stayed_nights	11
6	walking_distance	30	friendly_helpful	38	place_stay	10
7	stay_porto	29	value_money	38	location_perfect	9
8	highly_recommend	26	staff_helpful	35	stay_porto	9
9	great_location	25	city_centre	34	clean_comfortable	8
10	room_clean	23	friendly_staff	34	friendly_helpful	8

Table 4 - TripAdvisor Accommodation - top bigrams by nationality

The main results of this study are disclosed. In the next section, some conclusion are presented.



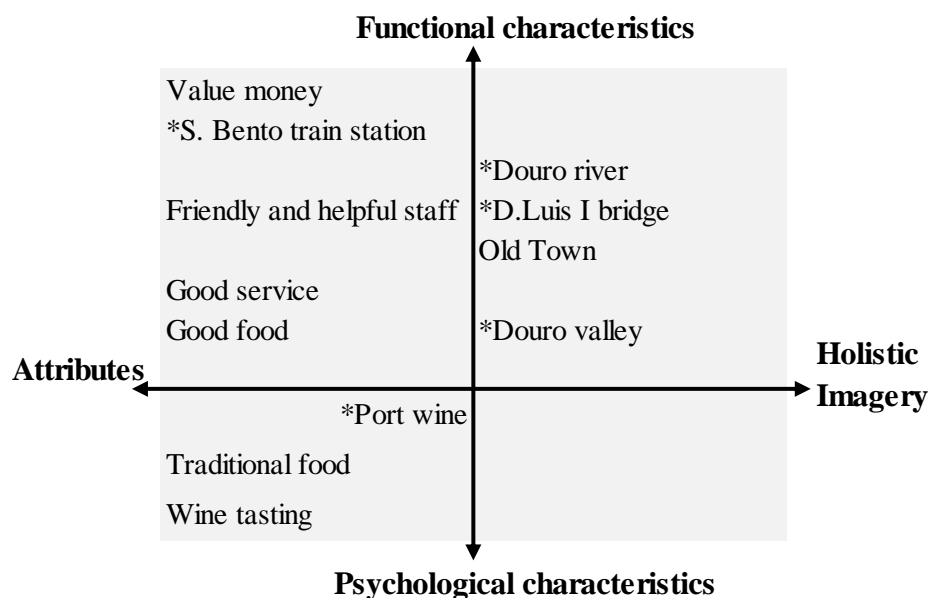
## 6. Conclusions

This is the first study about Destination Image developed for this region with this methodology. For that reason this investigation acts more as an experimental research, it intends to contribute with some insights of tourist's image of Porto but also encourage future researches to continue this project for deeper conclusions. In the next lines the main conclusions of this study are described.

Porto Wine and Douro River seems to be the central elements of Porto Image, being the most frequent bigrams observe in this research. When searching value judgements, beautiful city, great views, staff friendly and food good are the highlighted descriptions. Regarding the analysis comparing the most frequent bigrams by nationality, results suggest there are no significant differences. Though for conclusive answers, the study would have to examine a bigger sample and consider reviews in different languages.

It is possible to make some association of the main results with destination image components presented in literature review. Through a general overview of the highlighted bigrams, some suggestions are now exposed (figure 6).

Figure 6 – Destination Image components – adaptation to Porto



Concerning functional attributes, price levels and monuments as S. Bento train station can be mentioned. Also good evaluation on service and food quality are considered functional attributes. In functional and holistic quadrant, Douro River, D. Luis I bridge

and Old town seem to represent an overall and more holistic image of Porto, apparently it can be a "mental picturing" of the city. Stepping to the psychological attributes field, traditional food, port wine and wine tasting were considered psychological attributes as seems to create a relation to local traditions and culture. Additionally, the bigrams presented with an asterisk (\*) are considered "unique" elements, S. Bento train station, Douro river, D. Luis I bridge and Port wine seem to present the city identity and tough they are incomparable. All the other bigrams are considered "common" as they may be compared with other destinations.

It is also interesting to compare the results from travelpod and tripadvisor. While in blogs the tourist is free to write what he consider to be worthy, in tripadvisor reviewers have to choose from a structured list of categories, so the review is focus on a particular episode, theme or experience pre-categorized. Some of the most repeated words are common in both platforms, but in some other cases it is not. While in Attractions category, Serralves museum is highlighted with five bigrams related to modern or contemporary art, in travelblog none is visible. Bloggers apparently preferred to highlight other aspects of their experience. When comparing the blog results with Restaurants and Accommodation the differences are bigger. As previously mentioned, bloggers did not focus their speech in this two themes.

The study presented in section 2, where in graphic 2 are the words/expressions more voted by tourists when describing Portugal also illustrates some different results from the ones here presented. In that study climate and beaches are highlighted. Though, it is possible to find some similarities with hospitality, country beauty, good wines and good prices. Apparently Porto and Portugal may have different images. The different methodology may also help to explain the results, that research was made thru surveys.

In this Travel 2.0 reality, it is essential that tourism businesses take advantage of the different platforms offered by the new ICT to enhance their reputation. Among the various actions that companies should take in light of this new web generation are listening to customer feedback, building a direct relationship with them to determine their needs and respond to all kind of feedback (Muñoz et al., 2012).

## 6.1. Contributions

This paper has made a contribution to the analysis of the tourism sector in Porto. The conclusions of this paper may assist policy makers and private entrepreneurs on the formulation of their marketing strategy. In addition, it gives both public and private players the opportunity to review and improve their image positioning, to better attract and increase the potential of Porto as a tourism destination.

## 6.2. Limitations and future recommendations

In this study, only reviews written in English were considered and that limited the access to tourists' reviews from countries that may prefer to write in a different language. According to Intercampus (2013b), the countries that more visited Porto in 2012 were Spain (26%), France (14.3%) and Brazil (10%). In this study, the weight of those nationalities is much lower and so those opinions may be overlooked.

The results of this research may be influenced by an unbalanced number of reviews in each category in TripAdvisor. For instance, one Hotel may have 350 reviews and one other have just one, this last Hotel will be almost unrepresented. The same happens with individual reviews, if one comment frequently repeat one expression it may affect the global results.

As future recommendations we suggest a comparison of Porto image insights presented in this studies with the data presented and transmitted by official organizations or from the media press.

## 7. Bibliography

- Andrade, D. T. N. d. (2013). *Portugal sob o olhar francês - A imagem de destino turístico comentada online*. Faculdade de Economia do Porto.
- Andreu, L., Bigné, J. E., & Cooper, C. (2001). Projected and Perceived Image of Spain as a Tourist Destination for British Travellers. *Journal of Travel & Tourism Marketing*, 9(4), 47-67. doi: 10.1300/J073v09n04\_03
- Baloglu, S., & McCleary, K. W. (1999). A model of Destination Image formation. *Annals of Tourism Research*, 26(4), 868-897.
- Barich, H., & Kotler, P. (1991). A framework for marketing image management. *Sloan management review*, 32(2), 94.
- Boyd, D. m., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230. doi: 10.1111/j.1083-6101.2007.00393.x
- Buhalis, D., & Jun, S. H. (2011). *E-Tourism - Contemporary Tourism Reviews*. International Centre for Tourism and Hospitality Research Tourism Management and Marketing School of Services Management.
- Çakmak, E., & Isaac, R. K. (2012). What destination marketers can learn from their visitors' blogs: An image analysis of Bethlehem, Palestine. *Journal of Destination Marketing & Management*, 1(1-2), 124-133. doi: 10.1016/j.jdmm.2012.09.004
- Cheong, H. J., & Morrison, M. A. (2008). Consumers' Reliance on Product Information and Recommendations Found in UGC. *Journal of Interactive Advertising*, 8(2), 38-49.
- Cheung, C. M. K., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems*, 54(1), 461-470. doi: 10.1016/j.dss.2012.06.008
- Choi, S., Lehto, X. Y., & Morrison, A. M. (2007). Destination image representation on the web: Content analysis of Macau travel related websites. *Tourism Management*, 28(1), 118-129. doi: 10.1016/j.tourman.2006.03.002
- comScore. (2013). Media Metrix for TripAdvisor Sites, worldwide, Q1 2014. Retrieved 15 August 2014, from [http://www.tripadvisor.com/PressCenter-c4-Fact\\_Sheet.html](http://www.tripadvisor.com/PressCenter-c4-Fact_Sheet.html)
- Conrady, R., & Buck, M. (2007). Trends and issues in global tourism - 2007.
- Davydova, E. (2012). *The role of Travel Blogs in the process of Travel Planning*. (Bachelor).

- Dellarocas, C. (2003). The digitalization of WOM: Promises and challenges of online feedback mechanisms *Management science*, 49(10), 1407-1424.
- Do-Hyung, P., Lee, J., & Han, I. (2007). The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement. *International Journal of Electronic Commerce / Summer 2007*, V, 11, No. 4, 125-148. doi: 10.2753/JEC1086-4415110405
- Dutton, W., & Helsper, E. J. (2007). The internet in Britain 2007: Oxford Internet Institute, University of Oxford.
- Dwivedi, M. (2009). Online destination image of India: a consumer based perspective. *International Journal of Contemporary Hospitality Management*, 21(2), 226-232. doi: 10.1108/09596110910935714
- Echtner, C. M., & Ritchie, J. R. B. (1993). The Meaning and Measurement of Destination Image. *Journal of Travel Research*, 31(4), 3-13.
- Gallarza, M. G., Saura, I. G., & Garcia, H. C. (2002). Destination image towards a conceptual framework. *Annals of Tourism Research*, 29(1), 56-78.
- Google. (2013a). 2013-traveler\_research-studies. Retrieved 18 August 2014, from <http://www.thinkwithgoogle.com/research-studies/2013-traveler.html>
- Goyette, I., Ricard, L., Bergeron, J., & Marticotte, F. (2010). e-WOM Scale: word-of-mouth measurement scale for e-services context. *Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration*, 27(1), 5-23. doi: 10.1002/cjas.129
- Gretzel, U. (2007). *Online Travel Review Report*.
- Gretzel, U., & Yoo, K. H. (2008). Use and Impact of Online Travel Reviews. *Information and communication technologies in tourism 2008*, 35-46.
- Hofmann, M., & Klinkenberg, R. (Eds.). (2013). RapidMiner: Data Mining Use Cases and Business Analytics Applications. CRC Press.
- Hosany, S., Ekinci, Y., & Uysal, M. (2006). Destination image and destination personality: An application of branding theories to tourism places. *Journal of Business Research*, 59(5), 638-642. doi: 10.1016/j.jbusres.2006.01.001
- Intercampus. (2013a). Estudo Satisfação de Turistas: Análise de Resultados. Retrieved 01 Maio 2014, from [http://www.turismodeportugal.pt/Portugu%C3%AAs/ProTurismo/estat%C3%AAdsticas/EstudosdeSatisfacaodosTuristas/Documents/An%C3%A1lise\\_Satisfa%C3%A7%C3%A3o%20Turistas%202013%20-%20Vaga%20Ver%C3%A3o%20vf1.pdf](http://www.turismodeportugal.pt/Portugu%C3%AAs/ProTurismo/estat%C3%AAdsticas/EstudosdeSatisfacaodosTuristas/Documents/An%C3%A1lise_Satisfa%C3%A7%C3%A3o%20Turistas%202013%20-%20Vaga%20Ver%C3%A3o%20vf1.pdf)
- Intercampus. (2013b). Statistics, number of guests that stay at least one night in a touristic accommodation service.

- IpsosMediaCT. (2013). Google Travel Study. Retrieved 1 July 2014, from <http://www.thinkwithgoogle.com/research-studies/2013-traveler.html>
- Jalilvand, & Samiei. (2012). The impact of electronic word of mouth on a tourism destination choice: Testing the theory of planned behavior (TPB). *Internet Research*, 22(5), 591-612. doi: 10.1108/10662241211271563
- Jenkins, O. H. (1999). Understanding and Measuring Tourist Destination Images. *International Journal of Tourism Research*, 1(1), 1-15.
- Kantardzic, M. (2011). "Data-Mining Concepts." Data Mining: Concepts, Models, Methods, and Algorithms, Second Edition (2011): 1-25.
- Kesić, T., & Pavlić, I. (2011). *Tourism Destination Image Formation - The case of Dubrovnik, Croatia*.
- Kim, S., & Yoon, Y. (2003). The Hierarchical Effects of Affective and Cognitive Components on Tourism Destination Image. *Journal of Travel & Tourism Marketing*, 14(2), 1-22. doi: 10.1300/J073v14n02\_01
- Kotler, & Gertner. (2012). Country as brand, product, and beyond: A place marketing and brand management perspective. *Henry Stewart publications*, 1350-231IX *Brand Management Vol.9 No.4-5 249-261 April 2002*.
- Lai, L. S. L., & Turban, E. (2008). Groups Formation and Operations in the Web 2.0 Environment and Social Networks. *Group Decision and Negotiation*, 17(5), 387-402. doi: 10.1007/s10726-008-9113-2
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458-468. doi: 10.1016/j.tourman.2007.05.011
- LonelyPlanet. (2013). Retrieved 01 August 2014, from <http://www.lonelyplanet.com/europe/travel-tips-and-articles/77757>
- MacKinnon, K. A. (2012). User Generated Content vs. Advertising: Do Consumers Trust the Word of Others Over Advertisers? . *The Elon Journal of Undergraduate Research in Communications*, Vol. 3, No. 1.
- Martin, D., Woodside, A. G., & Dehuang, N. (2007). Etic interpreting of naïve subjective personal introspections of tourism behavior: Analyzing visitors' stories about experiencing Mumbai, Seoul, Singapore, and Tokyo. *International Journal of Culture, Tourism and Hospitality Research*, 1(1), 14-44. doi: 10.1108/17506180710729592
- Martineau, P. ( 1958). The personality of the retail store: 47-55.
- Mckee, R. (1997). *Story: substance, structure, style and the principles of screenwriting*. New York: HarperCollins.

- McKinsey, I. (2011). Big data: The next frontier for innovation, competition and productivity. Retrieved 10 September, 2014, from [http://www.t-systems.com/solutions/download-mckinsey-quarterly-/1148544\\_1/blobBinary/Study-McKinsey-Big-data.pdf](http://www.t-systems.com/solutions/download-mckinsey-quarterly-/1148544_1/blobBinary/Study-McKinsey-Big-data.pdf)
- Miguéns, J., Baggio, R., & Costa, C. (2008). *Social media and Tourism Destinations\_Tripadvisor case study* Advances in Tourism Research, (Aveiro).
- Muñoz, F., Hernández, J., & Sánchez, J. (2012). Generalising user behaviour in online travel sites through the Travel 2.0 website acceptance model. *Online Information Review*, 36(6), 879-902. doi: 10.1108/14684521211287945
- Nielsen. (2009). Nielsen Global Online Consumer Survey. Retrieved 8 July, 2014, from [http://www.nielsen.com/content/dam/corporate/us/en/newswire/uploads/2009/07/pr\\_global-study\\_07709.pdf](http://www.nielsen.com/content/dam/corporate/us/en/newswire/uploads/2009/07/pr_global-study_07709.pdf)
- O'Connor, P. (2008). User-Generated Content and Travel\_A case study on Tripadvisor. *Information and communication technologies in tourism 2008*, 47-58.
- O'Connor, P., Höpken, W., & Gretzel, U. (2008). Information and Communication Technologies in Tourism Book.
- O'Leary, S. (2005). Ireland's Image as a Tourism Destination in France: Attribute Importance and Performance. *Journal of Travel Research*, 43(3), 247-256. doi: 10.1177/0047287504272025
- O'Reilly, T. (2005). What Is Web 2.0 - Design Patterns and Business Models for the Next Generation of Software. Retrieved 15 September 2014, from [http://mediaedu.typepad.com/info\\_society/files/web2.pdf](http://mediaedu.typepad.com/info_society/files/web2.pdf)
- OECD. (2014). Tourism trends and policy priorities. Retrieved 25 September 2014, from <http://dx.doi.org/10.1787/tour-2014-4-en>
- Papathanassis, A., & Knolle, F. (2011). Exploring the adoption and processing of online holiday reviews: A grounded theory approach. *Tourism Management*, 32(2), 215-224. doi: 10.1016/j.tourman.2009.12.005
- Pike, S. (2002). Destination image analysis\_a review of 142 papers from 1973 to 2000. *Tourism Management*, 23(5), 541-549.
- Prats, L., & Marín, J. (2012). Blogtrip - Incostabrava or the use of bloggers as a destination image ambassador. *International Journal of Management Cases*, 14(4), 297-307.
- PubliTuris. (2013). from <http://www.publituris.pt/2013/12/06/porto-e-distinguido-pela-tripadvisor/>
- Quilkin, S. M. (2013). The Destination Image of Portugal from the Viewpoint of British Tourists: An Analysis of Online Blogs and Reviews.

- Reynolds, W. H. (1965). The role of the consumer in image building. *California management review*, 7(3).
- Reza, Samiei, Dini, & Yaghoubi. (2012). Examining the structural relationships of electronic word of mouth, destination image, tourist attitude toward destination and travel intention: An integrated approach. *Journal of Destination Marketing & Management*, 1(1-2), 134-143. doi: 10.1016/j.jdmm.2012.10.001
- Rong, J., Vu, H. Q., Law, R., & Li, G. (2012). A behavioral analysis of web sharers and browsers in Hong Kong using targeted association rule mining. *Tourism Management*, 33(4), 731-740. doi: 10.1016/j.tourman.2011.08.006
- San Martín, H., & Rodríguez del Bosque, I. A. (2008). Exploring the cognitive–affective nature of destination image and the role of psychological factors in its formation. *Tourism Management*, 29(2), 263-277. doi: 10.1016/j.tourman.2007.03.012
- Simms, A. (2012). *Online user-generated content for travel planning - different for different kinds of trips?* University of Wollongong.
- Son, A. (2011). International tourists' image of Zhangjiajie, China: content analysis of travel blogs. *International Journal of Culture, Tourism and Hospitality Research*, 5(3), 306-315. doi: 10.1108/17506181111156998
- Tan, A.-H. (1999). "Text mining: The state of the art and the challenges." Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases. Vol. 8.
- Thevenot, G. (2007). Blogging as a social media. *Tourism and Hospitality Research*, 7(3-4), 287-289. doi: 10.1057/palgrave.thr.6050062
- Tripadvisor. (2014). Retrieved 1 September 2014, from [http://www.tripadvisor.com/PressCenter-c4-Fact\\_Sheet.html](http://www.tripadvisor.com/PressCenter-c4-Fact_Sheet.html)
- Walzem, S. J. (2011). An American Tourist in Portugal.
- Wenger, A. (2008). Analysis of travel bloggers' characteristics and their communication about Austria as a tourism destination. *Journal of Vacation Marketing*, 14(2), 169-176. doi: 10.1177/1356766707087525
- Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. *Tourism Management Perspectives*, 10, 27-36. doi: 10.1016/j.tmp.2014.01.001



# Annex 1 – Rapid Miner

Stage 1:

**Edit Parameter List: crawling rules**  
 Edit Parameter List: **crawling rules**  
 Specifies a set of rules that determine, which links to follow and which pages to process.

rule application	rule value
follow_link_with_matching_url	http://www.travelpod.com/blogs/[0-9]+/Portugal/Porto.html
store_with_matching_url	http://www.travelpod.com/blogs/[0-9]+/Portugal/Porto.html

Specifies the value of the rule (string)

Stage 2:

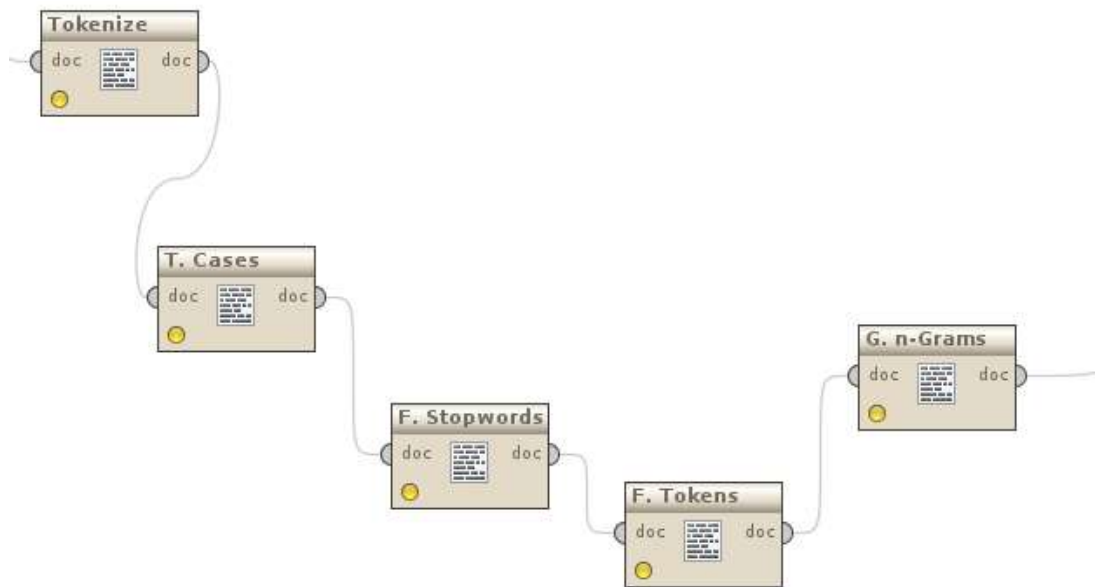
**Edit Parameter List: xpath queries**  
 Edit Parameter List: **xpath queries**  
 Specifies a list of attribute names and their corresponding XPath queries. See the operator documentation for details on XPath.

attribute name	query expression
title	//h1[div[@id='REVIEWS']/h2[div[@name
country	//h1[div[@id='REVIEWS']/h2[div[@class='location']/text()
comment	//h1[div[@id='REVIEWS']/h2[div[@property='description']/text()
name	//span[@property='name']/text()
Recommendation	substring-before(//h1[div[@id='PAGEHEADING']/h2[div[@class='rate sprite-rating_no rating_no']/h3[div[@class='col2of2_composite']/h4[div[@class='line cl
Excellent	substring-before(substring-after(//h1[div[@class='col2of2_composite']/h2[div[@class='line cl
VeryGood	substring-before(substring-after(//h1[div[@class='col2of2_composite']/h2[div[@class='line cl
Average	substring-before(substring-after(//h1[div[@class='col2of2_composite']/h2[div[@class='line cl
Poor	substring-before(substring-after(//h1[div[@class='col2of2_composite']/h2[div[@class='line cl
Terrible	substring-before(substring-after(//h1[div[@class='col2of2_composite']/h2[div[@class='line cl
SleepQuality	substring-before(//h1[div[@id='SUMMARYBOX']/h2[div[@class='col2of2_composite']/h3[div[@class='line cl
Location	substring-before(//h1[div[@id='SUMMARYBOX']/h2[div[@class='col2of2_composite']/h3[div[@class='line cl
Rooms	substring-before(//h1[div[@id='SUMMARYBOX']/h2[div[@class='col2of2_composite']/h3[div[@class='line cl
Service	substring-before(//h1[div[@id='SUMMARYBOX']/h2[div[@class='col2of2_composite']/h3[div[@class='line cl
commentSummary	//h1[div[@id='PAGEHEADING']/h2[div[@class='line cl
Value	substring-before(//h1[div[@id='SUMMARYBOX']/h2[div[@class='col2of2_composite']/h3[div[@class='line cl
Cleanliness	substring-before(//h1[div[@id='SUMMARYBOX']/h2[div[@class='col2of2_composite']/h3[div[@class='line cl
OriginalRecommendation	substring-before(//h1[div[@id='PAGEHEADING']/h2[div[@class='line cl
class	substring-before(//h1[div[@id='SUMMARYBOX']/h2[div[@class='col2of2_composite']/h3[div[@class='line cl

Specifies the name of the attribute to store the XPath query result in (string)

Add Entry Remove Entry Apply Cancel

Stage 3:



Stage 4:

☒ add meta information

☒ keep text

prune method none

datamanagement double\_sparse\_array

☒ select attributes and weights

**specify weights** Edit List (1)...

Stage 5:

```

1
2
3
4 x<-as.data.frame(data[c(1,3)])
5
6 require(wordcloud)
7 require(RColorBrewer)
8
9 countryVal<-as.data.frame(country)
10 filename<-paste(countryVal$país,".pdf", sep=" ")
11
12 pdf(filename)
13 wordcloud(x$word,x$total, scale=c(8,2),min.freq=2, max.words=Inf, random.order=FALSE, rot.per=.15, colors=brewer.pal(8, "Dark2"))
14 dev.off()
  
```

## Annex 2 - Blog data

Table bellow presents top 100 frequency words and bigrams.

	Total Words Freq.		Total Bigrams Freq.	
1	porto	1348	port_wine	146
2	port	833	douro_river	90
3	city	660	side_river	47
4	day	606	old_town	44
5	river	586	river_douro	43
6	time	539	white_port	43
7	went	531	douro_valley	40
8	got	501	nova_gaia	38
9	portugal	491	train_station	38
10	get	489	porto_portugal	34
11	wine	443	arrived_porto	33
12	town	426	day_porto	31
13	good	414	port_tasting	31
14	old	397	vila_nova	31
15	bus	396	year_old	31
16	see	382	city_porto	26
17	tour	378	wine_tasting	26
18	night	339	port_cellars	25
19	nice	337	atlantic_ocean	24
20	people	330	tour_guide	24
21	took	302	couple_days	22
22	made	291	decided_take	22
23	place	276	luis_bridge	22
24	great	267	city_centre	21
25	bit	266	harry_potter	21
26	portuguese	261	port_houses	21
27	walked	261	northern_portugal	20
28	found	260	port_caves	20
29	douro	256	river_cruise	20
30	hostel	256	world_heritage	20
31	hotel	245	beautiful_city	19
32	going	242	boat_ride	19
33	take	242	bus_tour	19
34	decided	239	city_center	19

35	dinner	227	dom_luis	19
36	find	226	hop_hop	19
37	beautiful	224	old_city	19
38	beach	218	porto_port	19
39	walk	217	san_sebastian	19
40	headed	216	bus_station	18
41	side	208	old_buildings	18
42	pretty	194	stock_exchange	18
43	streets	192	views_city	18
44	morning	191	good_thing	17
45	lisbon	188	great_time	17
46	train	185	half_hour	17
47	days	180	part_town	17
48	bridge	177	porto_city	17
49	arrived	175	pretty_good	17
50	lunch	174	ruby_port	17
51	hours	170	spent_day	17
52	trip	170	spent_time	17
53	couple	167	walking_tour	17
54	area	165	blue_white	16
55	room	165	bus_stop	16
56	food	162	caught_bus	16
57	walking	161	decided_head	16
58	know	158	port_lodges	16
59	think	156	port_porto	16
60	came	153	rio_douro	16
61	hour	153	went_walk	16
62	look	153	bus_porto	15
63	make	153	day_went	15
64	called	151	largest_city	15
65	lot	151	melted_cheese	15
66	saw	151	place_called	15
67	white	151	port_made	15
68	tasting	150	porto_river	15
69	looking	149	river_porto	15
70	stop	149	vintage_port	15
71	port_wine	146	wine_port	15
72	restaurant	146	bus_ride	14
73	street	146	city_portugal	14
74	top	146	got_bus	14
75	amazing	145	hop_bus	14
76	met	145	ice_cream	14

77	buildings	144	narrow_streets	14
78	thing	144	ponte_luis	14
79	local	142	port_tour	14
80	minutes	141	porto_wine	14
81	big	138	said_goodbye	14
82	church	137	south_side	14
83	free	136	tourist_office	14
84	said	136	town_porto	14
85	water	136	trip_porto	14
86	road	135	trying_get	14
87	things	135	bottle_wine	13
88	come	134	casa_musica	13
89	full	134	day_trip	13
90	car	129	finally_got	13
91	started	129	glass_port	13
92	looked	127	headed_hostel	13
93	told	127	heritage_site	13
94	flight	125	needless_say	13
95	thought	124	river_front	13
96	say	123	steep_hills	13
97	actually	121	streets_porto	13
98	spain	121	took_bus	13
99	spent	121	view_porto	13
100	amp	119	couple_hours	12

In table bellow are top frequency bigrams per country, those are de countries with more posts published in English.

	<b>Australia</b>	<b>Freq.</b>	<b>United States</b>	<b>Freq.</b>	<b>Canada</b>	<b>Freq.</b>
1	port_wine	31	port_wine	52	port_wine	25
2	old_town	20	douro_river	26	douro_river	22
3	train_station	19	arrived_porto	13	side_river	14
4	douro_river	18	porto_portugal	13	douro_valley	11
5	said_goodbye	14	san_sebastian	12	old_town	11
6	port_tasting	13	city_porto	11	train_station	11
7	white_port	12	side_river	10	year_old	11
8	bus_station	11	boat_ride	9	day_porto	10
9	day_porto	11	douro_valley	9	white_port	10
10	headed_hostel	11	pretty_good	9	nova_gaia	9
11	place_called	11	white_port	9	porto_portugal	9
12	went_walk	11	pedro_nuno	8	river_douro	9
13	couple_days	10	spent_day	8	river_valley	8
14	river_douro	10	wine_tasting	8	beautiful_city	7
15	side_river	10	atlantic_ocean	7	port_houses	7
16	arrived_porto	9	casa_musica	7	tour_guide	7
17	bus_stop	9	city_center	7	arrived_porto	6
18	harry_potter	9	city_portugal	7	atlantic_ocean	6
19	part_town	9	finally_got	7	bus_tour	6
20	port_caves	9	got_see	7	city_centre	6
21	went_hostel	9	morning_woke	7	glass_port	6
22	wine_tasting	9	port_cellars	7	ruby_port	6
23	world_heritage	9	year_old	7	vila_nova	6
24	bottle_wine	8	bad_news	6	yellow_bus	6
25	caught_bus	8	day_porto	6	aged_years	5
26	city_centre	8	decided_head	6	alcohol_content	5
27	douro_valley	8	decided_take	6	caught_bus	5
28	good_thing	8	famous_port	6	cost_euros	5
29	melted_cheese	8	hop_hop	6	day_took	5
30	old_city	8	largest_city	6	french_fries	5

## Annex 3 - Tripadvisor Attractions data

Table bellow presents top 100 frequency words and bigrams.

	<b>Total Words</b>	<b>Freq.</b>	<b>Total Bigrams</b>	<b>Freq.</b>
1	porto	1026	worth_visit	142
2	river	762	douro_river	119
3	visit	586	train_station	87
4	see	584	port_wine	77
5	beautiful	564	great_views	69
6	great	552	river_douro	60
7	bridge	531	visit_porto	49
8	worth	531	definitely_worth	48
9	city	478	see_porto	46
10	nice	462	great_view	45
11	view	418	nova_gaia	45
12	views	383	view_city	43
13	walk	378	city_porto	42
14	place	358	great_place	41
15	tour	355	view_porto	41
16	church	347	views_city	40
17	station	343	guided_tour	38
18	douro	340	side_river	37
19	get	336	luis_bridge	36
20	take	323	modern_art	36
21	top	313	walk_bridge	36
22	good	308	boat_trip	35
23	amazing	303	views_river	35
24	building	295	railway_station	34
25	time	280	sao_bento	34
26	day	266	place_visit	33
27	museum	258	vila_nova	33
28	port	245	casa_musica	32
29	area	232	view_river	32
30	old	224	wine_tasting	32
31	wine	220	cable_car	31
32	interesting	212	blue_white	30
33	train	209	porto_gaia	30
34	side	197	river_cruise	30
35	people	185	upper_level	30
36	tiles	185	worth_visiting	30

37	art	180	douro_valley	29
38	trip	180	worth_seeing	29
39	tower	176	art_deco	28
40	wonderful	168	boat_ride	28
41	restaurants	164	wine_cellars	28
42	cathedral	162	tile_work	27
43	boat	161	views_porto	27
44	gaia	161	take_time	26
45	look	158	art_museum	24
46	walking	153	contemporary_art	24
47	history	148	gaia_side	24
48	enjoy	145	old_town	24
49	lovely	144	see_beautiful	24
50	took	144	half_hour	23
51	worth_visit	142	top_bridge	23
52	gardens	140	visiting_porto	23
53	portugal	140	walk_river	23
54	free	134	amazing_view	22
55	portuguese	125	arab_room	22
56	room	122	entrance_fee	22
57	bit	119	nice_place	22
58	douro_river	119	ponte_luis	22
59	find	119	porto_city	22
60	impressive	119	rail_station	22
61	park	119	beautiful_place	21
62	excellent	118	nice_views	21
63	climb	117	top_tower	21
64	lot	117	tour_guide	20
65	guide	116	view_top	20
66	metro	116	visit_church	20
67	entrance	115	amazing_views	19
68	went	115	cross_bridge	19
69	architecture	113	day_night	19
70	euros	113	entrance_hall	19
71	work	110	palacio_bolsa	19
72	part	109	part_porto	19
73	night	108	porto_worth	19
74	cruise	107	river_side	19
75	lots	107	take_pictures	19
76	experience	106	beautiful_view	18
77	going	106	climb_top	18
78	ride	105	lower_level	18



79	town	103	nice_view	18
80	street	102	stock_exchange	18
81	fantastic	101	sunny_day	18
82	walls	101	take_walk	18
83	main	100	walk_top	18
84	bridges	99	good_place	17
85	level	98	nice_walk	17
86	make	98	port_houses	17
87	visited	98	view_bridge	17
88	visiting	98	boat_tour	16
89	especially	95	enjoy_view	16
90	miss	95	get_top	16
91	stunning	95	old_city	16
92	world	95	porto_take	16
93	definitely	94	take_look	16
94	modern	94	views_douro	16
95	shops	94	worth_effort	16
96	ticket	94	city_center	15
97	upper	94	eiffel_tower	15
98	minutes	91	good_views	15
99	narrow	89	see_city	15
100	hour	88	take_photos	15

In table below are top frequency bigrams per country, those are de countries with more posts published in English.

	<b>United States</b>	<b>Freq.</b>	<b>United Kingdom</b>	<b>Freq.</b>	<b>Canada</b>	<b>Freq.</b>
1	train_station	31	worth_visit	47	see_porto	5
2	douro_river	22	great_views	26	take_time	5
3	great_views	18	river_douro	17	worth_visit	5
4	worth_visit	16	side_river	15	tile_work	4
5	river_cruise	15	views_river	15	train_station	4
6	boat_ride	12	railway_station	14	view_city	4
7	river_douro	12	definitely_worth	12	clerigos_tower	3
8	walk_bridge	12	douro_river	12	douro_river	3
9	definitely_worth	11	port_wine	12	douro_valley	3
10	port_wine	11	views_city	12	gaia_side	3
11	casa_musica	9	worth_visiting	12	great_views	3
12	douro_valley	9	cable_car	11	port_wines	3
13	see_porto	9	blue_white	10	worth_stopping	3
14	views_city	9	entrance_hall	9	art_history	2
15	art_museum	8	half_hour	9	avoid_touristy	2
16	blue_white	8	luis_bridge	9	beautiful_view	2
17	view_city	8	walk_river	9	blue_tiles	2
18	wine_tasting	8	boat_trip	8	boat_ride	2
19	contemporary_art	7	good_views	8	cathedral_easy	2
20	see_beautiful	7	guided_tour	8	climb_top	2
21	see_city	7	port_houses	8	concert_meal	2
22	stock_exchange	7	top_bridge	8	corner_station	2
23	views_river	7	top_level	8	day_time	2
24	worth_stop	7	tour_guide	8	day_tour	2
25	art_deco	6	visit_porto	8	easy_find	2
26	great_view	6	wine_tasting	8	enjoy_view	2
27	sao_bento	6	boat_ride	7	entrance_hall	2
28	views_porto	6	great_view	7	free_entry	2
29	visit_porto	6	old_town	7	full_day	2
30	arab_room	5	views_douro	7	glass_wine	2

## Annex 4 - Tripadvisor Restaurants data

Table bellow presents top 100 frequency words and bigrams.

	Total Words	Freq.	Total Bigrams	Freq.
1	food	2464	food_good	252
2	good	2347	good_food	186
3	restaurant	2196	service_good	140
4	place	1574	good_service	118
5	porto	1531	good_value	118
6	great	1358	portuguese_food	117
7	service	1288	great_food	116
8	nice	1098	staff_friendly	112
9	wine	951	food_great	105
10	excellent	689	traditional_portuguese	92
11	staff	684	nice_place	91
12	menu	683	food_excellent	90
13	friendly	650	quality_food	90
14	portuguese	621	main_course	89
15	meal	595	value_money	86
16	dinner	527	friendly_staff	85
17	went	519	highly_recommend	79
18	delicious	512	excellent_service	76
19	time	472	great_place	75
20	atmosphere	471	wine_list	73
21	price	451	place_eat	70
22	night	444	food_service	67
23	lunch	443	excellent_food	63
24	eat	433	great_service	63
25	fish	424	friendly_helpful	62
26	quality	416	visit_porto	61
27	dishes	408	reasonably_priced	60
28	table	407	service_excellent	60
29	try	402	food_delicious	58
30	local	369	night_porto	57
31	recommend	368	service_great	56
32	people	355	food_nice	55
33	experience	351	food_wine	55
34	get	347	house_wine	53

35	recommended	345	friendly_service	52
36	find	343	highly_recommended	52
37	hotel	322	nice_restaurant	52
38	bit	321	red_wine	52
39	prices	316	service_friendly	52
40	day	315	restaurant_porto	51
41	waiter	311	bottle_wine	50
42	restaurants	306	good_quality	50
43	visit	305	went_restaurant	50
44	lovely	303	good_price	48
45	ordered	302	great_atmosphere	48
46	tasty	296	typical_portuguese	47
47	came	293	worth_visit	47
48	made	292	great_value	45
49	amazing	289	trip_advisor	45
50	worth	286	good_wine	44
51	evening	280	reasonable_prices	44
52	fresh	276	recommend_restaurant	44
53	portugal	274	service_food	44
54	served	272	port_wine	43
55	euros	270	restaurants_porto	43
56	river	270	good_place	42
57	main	263	portuguese_cuisine	42
58	value	262	definitely_recommend	41
59	dish	261	nice_atmosphere	41
60	traditional	259	restaurant_food	41
61	found	253	restaurant_good	41
62	food_good	252	good_choice	40
63	meat	247	restaurant_located	40
64	francesinha	241	douro_river	39
65	perfect	241	olive_oil	39
66	enjoyed	239	reasonable_price	39
67	view	234	sea_bass	39
68	port	233	white_wine	39
69	dessert	227	prices_reasonable	38
70	location	226	restaurant_nice	38
71	definitely	222	good_prices	37
72	sushi	221	good_restaurant	37
73	coffee	220	recommend_place	37
74	bread	218	restaurant_great	37
75	expensive	217	service_nice	37
76	reasonable	216	stay_porto	37

77	looking	214	found_place	36
78	helpful	211	great_experience	36
79	old	211	great_meal	35
80	ate	209	ice_cream	35
81	wonderful	207	glass_wine	34
82	city	206	nice_food	34
83	got	206	delicious_food	33
84	taste	202	staff_helpful	33
85	fantastic	201	cod_fish	32
86	beautiful	200	get_table	32
87	cheese	199	restaurant_recommended	32
88	choice	198	live_music	31
89	full	197	recommended_restaurant	31
90	want	195	family_run	30
91	salad	193	found_restaurant	30
92	course	192	local_food	30
93	tables	192	nice_service	30
94	tapas	190	place_great	30
95	good_food	186	recommended_hotel	30
96	located	186	vinho_verde	30
97	wines	183	food_amazing	29
98	come	182	great_location	29
99	house	180	lunch_time	29
100	street	180	place_good	29

In table bellow are top frequency bigrams per country, those are de countries with more posts published in English.

	United Kingdom	Freq.	United States	Freq.	Canada	Freq.
1	good_value	54	service_good	22	food_good	9
2	food_good	52	food_good	21	good_food	8
3	staff_friendly	31	highly_recommend	20	service_excellent	8
4	good_food	26	good_food	14	chocolate_cake	5
5	wine_list	24	food_excellent	13	food_delicious	5
6	main_course	23	good_service	13	food_excellent	5
7	quality_food	23	staff_friendly	13	glass_wine	5
8	service_good	23	house_wine	12	service_good	5
9	reasonably_priced	22	food_great	11	wine_list	5
10	night_porto	21	trip_advisor	11	excellent_food	4
11	traditional_portuguese	21	friendly_helpful	10	good_service	4
12	great_food	20	great_food	10	great_service	4
13	value_money	20	great_place	10	reasonably_priced	4
14	food_service	18	great_service	10	traditional_portuguese	4
15	food_delicious	17	olive_oil	10	ate_times	3
16	friendly_helpful	17	reasonably_priced	10	bottle_wine	3
17	good_service	17	staff_helpful	10	decided_give	3
18	great_place	17	stumbled_place	10	food_great	3
19	service_excellent	17	highly_recommended	9	glass_port	3
20	food_excellent	16	night_porto	9	highly_recommend	3
21	food_great	16	port_wine	9	language_barrier	3
22	food_wine	16	portuguese_food	9	night_porto	3
23	friendly_staff	15	reasonable_price	9	olive_oil	3
24	highly_recommend	15	red_wine	9	popular_locals	3
25	house_wine	15	wait_staff	9	portuguese_cuisine	3
26	good_choice	14	excellent_service	8	portuguese_food	3
27	service_friendly	14	food_service	8	restaurant_porto	3
28	bottle_wine	13	friendly_staff	8	service_great	3
29	excellent_food	13	ice_cream	8	andor_violeta	2
30	excellent_service	13	sea_bass	8	barrier_meal	2

## Annex 5 - Tripadvisor Accommodation data

Table bellow presents top 100 frequency words and bigrams.

	<b>Total words</b>	<b>Freq.</b>	<b>Total bigrams</b>	<b>Freq.</b>
1	room	2514	guest_house	312
2	hotel	2364	staff_friendly	256
3	porto	2079	friendly_helpful	224
4	good	1759	place_stay	215
5	staff	1589	stayed_nights	215
6	great	1582	walking_distance	203
7	breakfast	1490	metro_station	173
8	clean	1442	great_location	162
9	stay	1436	stay_porto	155
10	rooms	1371	good_location	150
11	stayed	1262	rooms_clean	141
12	hostel	1251	breakfast_good	138
13	nice	1208	staff_helpful	138
14	location	1198	value_money	138
15	place	1066	train_station	134
16	friendly	976	friendly_staff	130
17	city	965	room_clean	130
18	helpful	856	city_centre	125
19	night	802	city_center	116
20	nights	637	carmen_joao	110
21	walk	632	minute_walk	109
22	comfortable	625	minutes_walk	106
23	house	581	good_value	100
24	time	573	clean_comfortable	99
25	old	560	great_place	98
26	located	535	house_douro	96
27	bathroom	493	double_room	93
28	bed	492	helpful_staff	93
29	excellent	487	highly_recommend	91
30	station	483	breakfast_included	83
31	metro	475	helpful_friendly	80
32	perfect	468	location_perfect	78
33	day	466	location_great	77
34	recommend	464	sao_bento	77
35	price	463	stayed_hotel	76
36	get	458	visit_porto	76

37	street	456	hotel_located	75
38	people	430	top_floor	75
39	walking	426	hotel_good	74
40	area	423	spent_nights	74
41	river	402	good_breakfast	73
42	lovely	398	made_feel	73
43	beautiful	396	old_town	73
44	floor	382	great_value	71
45	service	381	clean_rooms	69
46	made	378	free_wifi	68
47	restaurants	372	definitely_stay	67
48	free	371	center_porto	66
49	wonderful	370	staff_nice	66
50	modern	355	star_hotel	66
51	guest	351	stayed_night	63
52	center	346	enjoyed_stay	62
53	centre	346	breakfast_room	59
54	view	346	location_good	59
55	amp	335	old_building	59
56	apartment	329	centre_porto	58
57	kitchen	328	front_desk	58
58	bit	327	nice_clean	58
59	amazing	321	centrally_located	57
60	definitely	319	port_wine	57
61	guest_house	312	glass_port	56
62	minutes	311	min_walk	55
63	building	309	definitely_recommend	54
64	visit	305	douro_river	54
65	find	304	clean_room	53
66	big	303	highly_recommended	53
67	town	300	central_location	52
68	make	294	hostel_stayed	52
69	quiet	293	room_spacious	50
70	value	292	extremely_helpful	49
71	spacious	290	hotel_staff	48
72	reception	288	night_stay	48
73	main	283	perfect_location	48
74	places	282	staff_extremely	47
75	port	272	air_conditioning	46
76	douro	271	city_porto	46
77	experience	270	clean_staff	46
78	see	270	make_stay	46



79	trip	269	room_nice	46
80	want	264	bed_comfortable	45
81	days	263	clean_modern	45
82	easy	262	excellent_location	45
83	beds	261	places_visit	45
84	feel	259	recommend_hotel	45
85	staff_friendly	256	room_bathroom	45
86	got	246	trip_porto	45
87	staying	243	bathroom_clean	44
88	shower	238	nice_breakfast	44
89	portugal	237	walk_river	44
90	bus	233	bento_train	43
91	arrived	230	casa_dos	43
92	food	229	porto_stay	43
93	loved	229	room_floor	43
94	double	228	hostel_staff	42
95	joao	228	breakfast_great	41
96	friendly_helpful	224	double_bed	41
97	distance	221	hostel_great	41
98	fantastic	218	hotel_great	41
99	top	216	nice_helpful	41
100	place_stay	215	orange_juice	41

In table bellow are top frequency bigrams per country, those are de countries with more posts published in English.

	<b>United States</b>	<b>Freq.</b>	<b>United Kingdom</b>	<b>Freq.</b>	<b>Canada</b>	<b>Freq.</b>
1	guest_house	84	stayed_nights	56	guest_house	37
2	carmen_joao	56	guest_house	48	house_douro	20
3	train_station	38	place_stay	47	walking_distance	15
4	house_douro	35	staff_friendly	47	carmen_joao	11
5	place_stay	33	metro_station	41	stayed_nights	11
6	walking_distance	30	friendly_helpful	38	place_stay	10
7	stay_porto	29	value_money	38	location_perfect	9
8	highly_recommend	26	staff_helpful	35	stay_porto	9
9	great_location	25	city_centre	34	clean_comfortable	8
10	room_clean	23	friendly_staff	34	friendly_helpful	8
11	stayed_nights	22	breakfast_good	33	highly_recommend	8
12	friendly_helpful	21	helpful_friendly	33	staff_friendly	8
13	made_feel	19	good_location	32	great_location	7
14	minute_walk	18	minute_walk	32	location_great	7
15	city_center	16	great_location	31	minute_walk	7
16	staff_friendly	16	clean_comfortable	30	room_clean	6
17	clean_comfortable	15	minutes_walk	30	staff_helpful	6
18	douro_river	15	room_clean	30	top_floor	6
19	front_desk	15	walking_distance	29	train_station	6
20	location_great	15	good_value	28	breakfast_good	5
21	breakfast_included	14	rooms_clean	28	breakfast_included	5
22	centrally_located	14	double_room	27	casa_dos	5
23	glass_port	14	stay_porto	27	definitely_stay	5
24	port_wine	14	old_town	20	front_desk	5
25	porto_stay	14	stayed_hotel	20	great_place	5
26	sao_bento	14	enjoyed_stay	19	metro_station	5
27	breakfast_morning	13	free_wifi	19	metro_stop	5
28	staff_helpful	13	air_conditioning	18	reasonably_priced	5
29	visit_porto	13	great_place	18	rooms_clean	5
30	definitely_stay	12	sao_bento	18	amazing_staff	4

