# QUANTITATIVE ANALYSIS
# OF HOTELS IN TRIPADVISOR:
# SPANISH AND PORTUGUESE DESTINATIONS

Author:
José Luis Ximénez de Sandoval Torres
(joseluis.xs@uma.es)


Co-Directors:
Antonio Fernández Morales
Sebastián Molinillo Jiménez


UNIVERSIDAD DE MALAGA


Málaga, (Spain)
June 2013

# TABLE OF CONTENTS

"A random walk is one in which future steps or direction can not be predicted from previous actions."

# 1 INTRODUCTION

Just as it is difficult to predict what will happen in the stock market tomorrow or in the coming month, a feeling of some random, uncontrollable risk has also affected the tourism sector.

Just as financial brokers and dealers sit at their computers early each morning to see how markets evolve, professionals in the hospitality industry start their day checking for new positive or negative comments on social networks that affect the reputation of their brands.
Up to now, marketing departments of companies have been controlling the image they wanted to communicate to society, creating advertising and promotion campaigns, controlling and negotiating with the media about what should or should not appear in the press.
A quote from Mr. Malkiel, professor at Princeton University, in his book "A Random Walk Down Wall Street" serves as a starting point for thinking about the revolution we are experiencing.

Just as the stock price is sensitive to any news appearing in the media about the company or its environment, the image of a brand that is being created or reinforced is also at risk of being damaged by a comment in social media on a daily basis.

The consumer, in this case, the tourist, has become the great authority, the prescriber, the advisor, the critic, the supreme judge, the expert, the friend, the peer, and whether the image and reputation that our Company has today remains the same tomorrow is in his hands. The future image of the company can not be predicted from past steps that have been taken by the marketing department. A simple online buzz can damage our brand very quickly.

Obviously all these changes would not have occurred without the development of new technologies. Web 2.0 has enabled consumer reviews which can be accessed by anybody with minimal effort.

If we talk about Tourism, Internet, web 2.0 and so on, it is necessary to mention the main player in this case: TripAdvisor, which will be one of the protagonists of our research.

Established in late 2000, it is the world's largest travel site. With over 100 million views and comments, over 200 million visits per month, every second it receives an opinion from tourists from around the world. Thus, it has become an invaluable tool for any tourist who is planning a trip.

With the importance of tourism in Spain, and because for some experts TripAdvisor has changed the rules of the game in the tourism industry, innumerable questions arise that need answering:

• What kind of comments do we find as we look through the hotel reviews on TripAdvisor?
• Is it a random sequence of unrelated comments or are there underlying patterns triggered by a variety of factors?, namely:
• The time of year: During summer months, are there a higher proportion of positive comments than in other seasons?
• The category of the hotel: comments about 5-Star hotels are more critical than of 4-Star hotels?
• The hotel: opinions of city hotels are more negative than those of beach hotels?
• The elements criticized: is the review about breakfast, location, clean rooms and attentive staff?
• Personal characteristics of the writer's comment: are romantic travelers more critical than families of travelers?, women more than men ?, young more than mature ?, domestic over

foreign?

• The pull factor: from a negative review, do more negative reviews or  positive reviews arise to offset one another?

• Being so easy to publish an opinion, what is the proportion of real to false comments?, and false positive to negative comments?

• To what extent do these opinions affect the image of the establishment?

• How do these comments or the position at which an establishment is ranked influence the level of employment and salaries?

• Finally, as a result of the above, are hoteliers aware of what some experts have called "The Tripadvisor Effect"?

We aim to answer these questions and some more as we are facing a new way of doing things in tourism:

• A new way to interact with customers and companies.

• A new balance of power.

• A new way for customers to learn to make decisions because now it is easier to learn, which paradoxically makes it more difficult to make decisions, having too much information available at a very low cost.

• A new way for companies to be informed and make decisions about their communication strategies, positioning and brand image on performance appraisal, motivation and compensation of personnel. Now knowing what the customers think about our brands, our products or services is easy, cheap and even instantaneous.

This revolution has only just begun. Are companies aware that the rules of the game have changed?  We do not live in times of change, but in changing times.

In any case, even if we have answers to this set of questions, it will not be easy with our outdated thought processes to understand how businesses operate in the 2.0 era.

- Can a company make millions of dollars with no more than comments written on its website for free by the clients in restaurants and hotels?
- Can these reviews, written by neither professional critics or prestigious writers, but simply customers who do not receive material compensation of any kind, damage the image of a hotel brand built up as a result of numerous million-dollar advertising campaigns?
- When making our purchasing decisions, can we rely on comments written by people who we do not know, with whom we have never spoken or whose tastes, lifestyle, real names are unknown to us?

Being aware that these last three questions are answered affirmatively, we must recognize that, although we are not able to understand it, although we do not believe it, everything has changed, everything is different, everything is new.

# 2 JUSTIFICATION OF THIS PAPER

As experts and tourism professionals noted, customer opinions, reviews, comments and ratings of hotels included in web sites like TripAdvisor have increasing influence on the decision making process of tourists. Potential customers rely more on the reviews written by others than on the recommendations found in official destination pages or on hotel websites, transforming the traditional WOM (word of mouth) into the new e-WOM (electronic word of mouth).

The importance such websites have acquired for the tourism sector has caused not only professionals, but also the research community to analyze the information contained in them, finding in recent years extensive literature from different points of view discussing this phenomenon.

Moreover, sun and beach tourism remains the fundamental axis of the tourism industry in countries like Spain and Portugal, and in an environment of economic crisis these countries are relying on this sector as a driver of recovery.

From these two major areas, (i) coastal tourism and (ii) assessment of hotels in the web 2.0, our work focuses on the analysis of the scores posted on TripAdvisor for hotel customers in sun and beach destinations in Spain and Portugal.

The goals of our work are:

1- To review the scientific literature on the creation of online content by users in the tourism sector.

2- To determine whether there is any relationship between some of the most important criteria of a hotel: Number of Reviews, Hotel size, Average rating. This analysis is performed globally (without segmenting the data by destination).

3- To compare the key indicators listed in TripAdvisor for hotels in 14 major coastal tourist areas of Spain and Portugal.

4- To compare the key indicators listed in TripAdvisor for hotels in the cities on the Costa del Sol.

Finally, this paper is structured as follows:

In Chapter 3 we review the contributions of other authors regarding the online comments about tourism.
In Chapter 4 we review the most characteristic aspects of TripAdvisor, and we present a detailed account of the information available on this website, which will be the main source of information for our work because of it's reference page for comments about hotels in most countries.
In Chapter 5 we describe the methodology of quantitative analysis applied in our research.
In Chapter 6 we present the results of our work.
Finally Chapter 7 presents the main conclusions and the possible applications and future research.

Furthermore, our work includes Appendix 1 with examples of hotel reviews and Appendix 2 with tables and graphs illustrating the variables studied.

# 3 TOURISM AND SOCIAL NETWORKS

Social networks are playing an increasing role as a source of information for travelers (Gretzel & Xiang, 2010). More specifically, the user-generated content on the Internet is changing the way consumers buy products and services (Ghose et al., 2012). Today those planning a trip, have more confidence in reviews of other tourists than in the advice from a travel agency (Ong, 2012).
The scientific literature on the analysis of customer reviews on the web is characterized by (i) being very extensive, (ii) being highly interdisciplinary and (iii) being quite recent.

(i) There have been a lot of articles published in the last few years in prestigious journals on various topics related to the subject: User-Generated Content, Electronic Word-of-Mouth, Online Reviews, Social Media, Opinion Mining, Sentiment Analysis.
(ii) We find research related to these subjects in a variety of disciplines: Business Administration, Marketing, Psychology, Geography, Statistics, IT Studies and of course in Tourism.
(iii) Non-stop growing literature devoted to these topics. In fact, some journals have recently dedicated special issues to these items:
Journal of Travel & Tourism Marketing. Special Issue on social media. 2013

Cornell Hospitality Quarterly. Special Issue on Advances in Hospitality and Information Technology Strategy. 2013
Marketing Science. Special Issue on the Impact of User-Generated Content. 2012
Computers in Human Behavior. Special Issue on Web 2.0 in Travel and Tourism: Empowering and changing the role of travelers. 2011

We can classify the countless works related to tourists online comments and opinions focusing on the three elements involved in the process: (i) The Supplier, (ii) The Customer and (iii) The Technology.

Thus we find:

1- Studies from the point of view of the Seller (Companies and Destination)
2- Studies from the point of view of the Customer, or Tourist
3- Studies from the point of view of the Technology


## 3.1. STUDIES FROM THE POINT OF VIEW OF COMPANIES AND DESTINATIONS

We found lots of works focusing on studying the impact of social media on tourism but also in the organizations responsible for destination management.


### 3.1.1 The TripAdvisor effect.

Smyth et al. (2010) analyzed more than 30,000 hotel reviews in Ireland, and 50,000 comments from hotels in Las Vegas for two years, and compares all of them.
While the average score of the hotels in Las Vegas has remained constant around 3.8 points (range of 1 to 5), Irish hotels have an increase in the same two years from an average score of 3.6 to 3.8. Furthermore, 64% of Irish hotels with a previous score of between 2 and 3 points had a rating superior to 3 at the end of the study period. For authors this is due to what they call "The TripAdvisor Effect: " While hotels in Las Vegas had reached acceptable levels of quality a long time ago, Irish hoteliers, being aware that some negative comments on TripAdvisor can damage the image of their establishments, have worked to improve their levels of service and avoid future negative comments.

### 3.1.2. Related to income

How do the customer reviews affect income or occupancy? Luca (2010) poses this question to the restaurants in the city of Washington and the customer review on Yelp.com, (webpage of customer reviews of any kind of industry in the USA) reaching the following conclusion: An increase of one star rank in yelp.com increases sales between 5% and 9% in independent restaurants. Although this does not affect restaurants that belong to large chains, the market share of these restaurants has decreased, while the relevance of Yelp.com has been increasing. For authors, the online comments are replacing traditional forms of reputation.

As for hotels, the industry is aware that scores of consumers have a strong impact on the decision to book a hotel. When an establishment has negative comments Verma et al. (2012) found that there is a 40% chance of customers booking that hotel. When the comment is positive the probability goes up to 70-80%.

### 3.1.3. In relation to business management

The extent to which managers manage reviews about their hotels listed on TripAdvisor has been studied by Park & Allen, (2013), who note that hoteliers tend to respond to comments that are considered honest and reflect the views of their customers, while those who do not respond to comments act this way because they believe they represent only extreme points of view (positive or negative).

Levy et al. (2013) studied 1,946 very negative comments (those who gave the property the lowest score from 1 to 5) extracted from the top ten websites with reviews about hotels in Washington DC and how to answer them, finding that the most common issues are related to the reception, bathroom, room cleanliness and noise. Hotels with best scores tend to respond to online complaints, with apologies and explanations. It is unusual to use offsets for mistakes. The authors believe that the growing role of social networks require that hotels use the online feedback to conduct market research and identify opportunities to improve their services.

Melián et al. (2010) recommend the hotel managers to promote the participation of its clients since according to these authors, the fewer negative ratings get a hotel are the same.

## 3.1.4. Online reputation

An item that is generating a large number of studies is online reputation. How do comments affect the image of a company or destination?

The importance of feedback to the brand value has been studied by Callarisa-Fiol et al. (2012). For these authors tourist accommodation is a key to creating a positive image of the destination from the customer's perspective. Analyzing only the quantitative scores (no comments or qualitative evaluations) of a sample of 45,914 and 8,736 tourists who had commented on TripAdvisor about their stay in hotels in Paris and Hong Kong respectively, the authors note that: the first city is recommended by visitors to the second, interpreting this as a sign of greater satisfaction.
They also point out that stars have an important role in shaping the brand value and the number of stars is not a clearly justifiable factor in the price of hotels. Finally the authors indicate that customer loyalty and positive recommendations on the establishment only occur from sitting at home, they receive a good service that corresponds with the hotel category and appropriate management of stay by staff and their managers.

Kwok & Yu (2012) analyzed in a restaurant review that most successful in Facebook checking that the more commercial content marketing and the comment had less "likes" received. They have also studied social networks Jain et al. (2012) noted that only 29% of travelers who use social networks are a mark of the tourism industry in a social network.

## 3.1.5. The tourism destination

A special section deserves analysis from the point of view of the destination.

Johnson et al. (2012) study the benefits of generating these comments for a tourist destination in the Canadian province of Nova Scotia, from the comments in TravelReview. Introducing the destination name as a search term collected 4,064 reviews of hotels (71%), 1,513 Restaurant (26%) and 153 on attractions (3%) in the area. Grouping entries on a geographical basis, all 77 localities within the region of Nova Scotia, lets you analyze the spatial distribution of the comments and how these vary across the destination. The information contained in such websites may offer a tool for DMO (Destination Marketing Organization).

The influence of the reviews posted in TripAdvisor by tourists to a destination, is studied by Miguens et al. (2008) through comments about hotels in the city of Lisbon. The results show that

the Portuguese capital has few hotels (191) represented on TripAdvisor, considering this as a weakness for online marketing strategies. Finding a large number of written comments by local citizens is evidence for these authors of the interest of the people of Lisbon to welcome tourists and potential should be considered an important asset to the city, having encouraged this participation from local authorities. Finally the authors believe that this study is further confirmation of the importance that social networks may have to promote both individual operators and the destination as a whole.

Returning to social networks, Stankov et al. (2010) study on Facebook participation of European National Tourism Organizations (NTOs), indicating that this network offers many ways to promote tourism and therefore the agencies responsible for promoting destinations should take it as a tool for generation of content by users. The work of these authors is to identify and quantify how much use of this network the NTO makes, finding that at the time the research was done (June and July 2009) it did not have a significant presence on Facebook and almost half (51%) of these organizations did not use Facebook on those dates. This data shows how slowly they are responding to the opportunities offered by the most popular social network in the world.

The relationship between DMO (Destination Marketing Organization) and e-WOM is analyzed by Tham et al. (2013). DMOs need to invest time and effort to understand how the e-WOM can generate great influence and appeal for the destination among potential visitors from pre to post-trip stages. For these authors the DMOs should provide opportunities for visitors to discuss and transfer their experiences of social networks in the destination. Therefore the ability to collect these experiences and the DMO's availability to answer visitors help the destination to be seen in a more favorable view. The success of a destination depends on their ability to influence the choice of tourists. Finally, these authors know the strengths and weaknesses of e-WOM allow DMO's strategic positioning in the digital marketing needs for possible tourist in the decision making process of the destination.


## 3.2. STUDIES FROM THE POINT OF VIEW OF THE CUSTOMER

This work focuses on the figure of the customer, who can play a dual role:

1- As a generator of content. One who, after a stay or a trip, recounts his experiences writing reviews, sharing photos and videos, etc.

2- As a user of the content. Whoever in their decision making process to choose a destination and / or hotel for your next trip, searches and reads the comments that other people have written.


### 3.2.1. The customer who writes comments

These works seek to answer questions like why comments are written?, or how tourists who write such comments can be classified?.

Over 1.5 million travelers` comments were written in USA Hotels in 2011, 45% more than in 2010, with 2011 the second consecutive year of double-digit growth (Jain et al., 2012)

OTAs (Online Travel Agency) as Expedia or Booking remain the most popular websites where travelers write their comments, collecting in 2011 70% of the reviews written by travelers, leaving the other 30% for specific pages of comments. In this segment TripAdvisor has made significant efforts to encourage travelers to write reviews, and has become the only major site specializing in reviews and opinions on hotels and restaurants both in terms of visitor traffic and posted reviews.

In 2011 opinions grew 69% in International while the OTA's only 37% (Jain et al., 2012)

The factors that lead someone to be interested in a review or upload some pictures illustrating their experiences as tourists are analyzed from multiple viewpoints. Thus, Wilson et al. (2012) carried out a comparative study of the motivations for writing reviews based on the nationalities of the individuals, comparing the motivations of tourists from Spain, Switzerland and the UK, revealing that the British and Swiss are more likely to share photos or stories through social networking (Facebook, for example) while the Spanish prefer to comment on TripAdvisor or on the website of the supplier (hotel, airline, etc.), concluding that tourists from different countries have different motivations for sharing their experiences. Marchiori et al. (2011) also analyze reviews written by tourists from various countries (in this case 10 different nationalities) on 77 hotels in Switzerland, and also agree that the nationality of tourists influence the type of comment. It also shows that the valuations of the hotels on Tripadvisor and Booking.com are similar.

Stringam et al. (2010) analyzed through scores collected on the page of Expedia (an online travel agency that allows tourists to discuss their experiences in a similar fashion to TripAdvisor) the correlation between the overall assessment about the hotel and intend to recommend and assess four specific aspects (service, facilities, cleanliness and comfort in the rooms). Among the results, he highlights that tourists tend to give high marks to the four aspects mentioned above (at least 70% of responses gave a score of between 4 and 5). If the initial perception of hotel professionals and some researchers was that most of the reviews are written by disgruntled or dissatisfied customers, the results of this study present evidence to the contrary. The 60,648 comments on 10,537 hotels in the 100 largest cities in the USA show that the majority of travelers who write comments about a hotel would be willing to recommend it (74.51%), and in the same line, the 72.85% who gave the hotels a score between 4 and 5 out of 5 where only 13, 43% of travelers assigned a 1 or 2 in terms of overall satisfaction. As for consistency in scores, they tend to be consistent, since all variables are strongly correlated.

Factors influencing users to write reviews are studied by Jurca et al. (2010) for whom there is a clear dependence between the comments and ratings from other customers for the same hotel. Previous experiences of other people generate expectations about service quality that influence writing of the comments. In addition, the overall score of a hotel being an average of all individual ratings received by that establishment, the authors ask how to change these overall scores if instead, the arithmetic mean median or mode are used. Using median, ranking position differs average positions 7,7 and 16,9 mode using positions on the calculated arithmetic mean. Therefore various measures can completely change the list of establishments presented to the user. For these authors, the median and mode that provides more robust performance against outliers might intentionally distort the overall rating of a hotel.

These authors also analyze how the position of a hotel evolves over time in the rankings as new comments appear. As analyzed with a particular hotel in New York, the results clearly show that the median and mode provide much more stable than the arithmetic mean rankings. They conclude that the lists based on scores obtained through arithmetic induce the user to inflate your score to move in the desired direction as far as the overall hotel rating possible. However, if the median were to be used, the user would be interested in providing the most accurate assessment possible.

Bronner & De Hoog (2011) who analyzed a sample of 3,500 Dutch tourists, whose comments were posted, discussed how and what personal characteristics are can be applied to people under 55 and couples, with or without children. The motivating factors for writing a comment according to these authors would be; (i) personal satisfaction, (ii) assist other tourists, (iii) social, (iv) increase the power of consumers and (v) help businesses.

There is a difference between the number of users who read and use comments and the number of tourists who write them. Gretzel & Yoo (2010) wonder what motivates these tourists to write and what makes them different from the rest, concluding that personality is a distinguishing feature

between those who write comments and those who donot.

After examining a sample of 100 negative reviews posted on TripAdvisor, Vasquez (2011) finds that a large number of them combine positive and negative reviews, and that a significant proportion refers to unmet expectations.

Lee et al (2011) analyzed more than 900 comments in hotels throughout the world, discovering that the most helpful reviews for users seeking information are those who travel most, publish many comments, and give low scores. No differences in gender and age were noted.

## 3.2.2. The clients who read reviews

The works in this section try to analyze aspects of the process of decision making and the ability to influence the reviews.
A report by the travel web minube.com (2011) shows that the recommendation is the basis for decision-making for customers arriving to the hotels without intermediaries. In this sense indicates Jain et al. (2012) the voice of the traveler is playing an increasingly important role in the search process for the purchase of travel. Thus, 46% of adults between 18 and 34 use the Internet to find reviews of restaurants that have never been written (Ong, 2012)

For Casaló et al. (2011), the intention to follow the advice in a published review depends on at least of 3 groups of factors; (i) those related to the nature of the advice, (perceived usefulness of the advice) (ii) those related to the source that provides the council (confidence in the web page that provides the comments) and (iii) those related to personal characteristics of the traveler who decides to follow the advice or not.

Other authors compare consumer attitudes in their purchase decision process between restaurants and hotels. So Ong (2012) discusses how the user perceives the reviews on hotels and in restaurants because they understand that it is not just the influence of a comment that matters when choosing a hotel to choosing a restaurant, the latter being a hedonic choice.

An interesting approach is proposed by Deutch & Milo (2012) when discussing the Mob Data Sourcing, noting that the information and data generated by the crowds (Mob) democratize data collection and pose a revolution in the world of information.

Other studies analyze how they influence the characteristics of the trip (familiarity with the destination, location, distance to destination ...) in the way that the tourist uses online feedback to plan their trip. For Simms (2012), the characteristics of travel, such as if it is a first visit to the destination, if it is an international tourist destination or if the tourist is traveling alone, play an important role in the use of social networks.

The differences in behavior between men and women, are studied by Verma et al. (2012) through a survey of 2,830 hotel clients in leisure travel and business within the United States, indicating that the main factor in choosing a hotel for a business trip is the recommendation of the company or organization, while as for those traveling for leisure, the number one source of information for choosing a hotel is the recommendation of friends and family. According to him, women are significantly more likely to read TripAdvisor reviews compared to men. Moreover, the planning of the trip is, according to these authors, as follows: In the early stages search engines (Google, Yahoo, Bing) are used, to gather destination information. Then go to the official website of the brands or destinations and OTAs (Expedia, Orbitz, Travelocity ...). At this point travelers read reviews on sites like TripAdvisor (OTA's include in their pages with comments).

## 3.3. STUDIES FROM THE POINT OF VIEW OF TECHNOLOGY

From a historical perspective we are in a new era in relations between the (Mistilis & Buhalis, 2012) technology and tourism

There is a lot of work from a technological point of view to analyze the phenomenon of the comments and opinions online.

Placing ourselves within the scope of PLN (natural language processing), the research is directed towards any of the following sites

### 3.3.1 Ontologies, folksonomy and concept maps

Xiang et al, (2009) focus on the semantic representation of the tourism domain, comparing ontologies of tourists in the use of search engines, ontologies and information derived from the official sites of tourist destinations, indicating a small number of common words between the two ontologies. Buttinger et al. (2010) describe a model ontology for tourist prices. Kim et al. (2010) present a semantic representation for folksonomies.

### 3.3.2 Opinion mining and sentiment analysis

Moreno Ortiz et al. (2010) performed an analysis of sentiment on the reviews on TripAdvisor Sentitext using the tool.

Esuli and Sebastiani (2006) develop a Sentiwordnet tool, which allows you to associate a value to a text as having a positive, negative opinion or having no opinion.

Peñalver-Martínez et al (2011) define a new method for opinion mining based features, although in this case it is applied to reviews of movies.

Carmona and Carrasco (2011) designed a linguistic fuzzy sumarizador with application to tourism marketing.

Hu and Wu (2009) propose a model for extracting comments and summarize the pros and cons of each review.

Xu et al (2011) a good summarization, besides the traditional requirements of relevant aspects and intensity of feelings must meet two others: representativity and diversity

### 3.3.3. Summarizing

Carrasco y Carmona (2011) design a fuzzy logic linguistic summarizer with application in tourism marketing.

Hu y Wu (2009) propose a model to summarize the comments extracting Pros and Cons from every comment.

For Xu et al (2011) good summarization besides the traditional requirements of relevant aspects and intensity of feelings must meet two others: representation and diversity

## 3.3.4 Viewing

Bjorkelund et al. (2012) have developed a prototype to show the changes in feelings about a hotel establishment over time.
To represent changes in feelings about a graphic used www.flotcharts.org. Many of the graphs have large fluctuations that stabilize at first but fluctuate over time, this is due to the distribution of opinions. In the early stages there are few comments and one extreme view has a large impact on the average. These authors determine the subjectivity and feelings of a document is one thing, finding the general feeling in a large dataset is quite another. What one person thinks of a product is not normally of interest. 10,000 people who think the same about a product is almost always interesting.

Akiva et al. (2008) designed an algorithm (BAM: Brand Association Map tool) to display how consumers talk about brands online, monitoring millions of conversations.

## 3.3.5 Classification Systems

Ye et al. (2009) to the large amount of information we have to study the decisions of sentiment classification techniques to facilitate decision-making by comparing how various algorithms operate seven destinations in USA and Europe. Reaching a similar reliability in all cases (80%) provided it has a large number of views available.

Paroubek and Pak (2010), design a way to classify feelings for comments on Twitter.

## 3.3.6 Recommendation Systems

Popescu et al. (2009), design a recommendation system based on collaborative filtering to recommend tourist destinations, through the information contained in the annotations of the photos on Flickr.

O'Mahoney et al. (2009) describe a recommendation system using Hotel Reviews in TripAdvisor.

Linaza et al. (2011), describe an application that, by selecting a series of images by tourists, set your profile, recommending destinations best suited to it.

García-Crespo et al.(2011) present an expert system for recommending hotels (Sem-Fit). This

system is based on fuzzy logic techniques to link consumers with the characteristics of the hotel.

Montejo-Raez et al. (2011), They design a system that combines planning by the user with the web recommendation, one of its objectives when designing the tool interface is simplicity and ease of use.

### 3.3.7 Data Extraction

Although this issue is technically solved some papers collected solutions to extract data from the web. For example Stringam & Gerdes (2008) includes a pseudocode for designing a web spider

### 3.3.8 Readability and Helpfulness

Given the proliferation of user-generated content, the implementation of automated processes that help users to easily access the most relevant comments and with more quality content is needed. (O'Mahony et al. 2010).They propose the creation of interfaces to help write more useful and easy to read, alerting use too long sentences or complex words comments, and comparing the degree of readability with those considered most useful. Using machine learning techniques they could select only the most interesting to help in decision-making, automatically distinguishing the useful comments that do not contribute comments. They think that there is a great production of opinions by users but many of them are not of interest.

For example, 25% of comments about hotels in Chicago picked up at their sample do not receive feedback of any kind, and only 35% receive it on 5 or more occasions.

For these authors the most discriminating factors in terms of usefulness of a comment are: (i) the number of words, (ii) the number of complex words (more than 3 syllables) and (iii) the number of sentences in the comment.

## *3.4. OTHER STUDIES*

There are some jobs that analyze online reviews from multiple angles:

1- The process of searching for information,

2- The writing process

3- The impact on the establishment

One of the first studies on the impact of TripAdvisor was Gretzel (2007) which analyzes the process of planning the trip, and the influence of online reviews from other tourists and the reasons for writing a review after a trip. It also examines the characteristics of user opinions and writers` reviews.

O'Connor (2010) examines a sample of hotels in London to determine the factors that cause more satisfaction and dissatisfaction among customers. It also analyzes the attitude of managers towards the comments, pointing out that there are very few who manage its reputation on TripAdvisor, wondering if hotels which actually generated content users are taken seriously. Finally

he highlights the major features of false opinions.

The publication of comments provides a new channel of communication between property, customer and competitors. Brown (2012) studies how the sites` comments can be designed to serve a wide variety of interactions around opinions beyond simple recommendations.

As highlighted by the aforementioned report minube.com (2011), all experts agree that the Internet has changed communication habits and consumption. In tourism, Internet has radically changed the entire process from making the decision to visiting a destination, to ways of sharing the experience with others. Technology has influenced everything including the information available to the user. The new traveler who has emerged from this digital revolution has caused the tourism sector the demand for price or terms have been replaced by others, such as communication, interactivity, dialogue, emotion, experience and inspiration. Travellers have gained visibility and independence in decision making and that has forced the industry to react.

We wanted to leave the hard work of cataloging until the end that, although it diverts somewhat from the core of our research, deserves to be taken into consideration by the originality with which it discussed the topic of online comments. These authors (Orlikowski & Scott, 2012) although the ranking mechanisms are not new, have acquired great influence and power of capacity when run through Web 2.0 technologies, redistributing the responsibility. The study focuses on small business remote unidentified geographic area for which TripAdvisor has changed the rules of the game. The authors analyze the moral and strategic implications of this transformation. These authors have been particularly interested in examining the concept of responsibility (accountability) applied to social networks. The concept of online liability is based on the idea of the wisdom of the crowd "wisdom of crowd" and the collective intelligence of "collective intelligence" produced by social media (Benkler 2006 (Surowiecki 2004).

The power of this ranking in its ability to present itself as an objective tool without bias, showing the truth, imposing their way of seeing things without being seen (its way of seeing, without being seen)

TripAdvisor has a material effect on the business and management. In fact, for certain tourism businesses, a better or worse score in the ranking can mean the difference between profit or loss, surviving a season or having to close. Therefore it is necessary to take the accountability of relationships that occur in TripAdvisor seriously, and the result is in practice a new kind of authority that directs our lives.

Finally highlight the paradox that TripAdvisor is configured in such a way that produces a form of transparency for comments from tourists while obscuring the mechanism that directs these comments.


## 3.5. FRAUD DETECTION

A special section of the text-mining, widely studied but still not resolved due to the difficulties and challenges is related to the detection of false or fraudulent comments (Ott et al., 2012)

Ong (2012) questioned whether the amount of feedback affects the perceived reliability thereof. Indeed a large number of reviews provides credibility. Perhaps users are aware of the possibility that some comments are false, but when the number of reviews is high (in the hundreds) then the impact of the manipulation of a fake review drops dramatically.

Wu et al (20110) present a set of criteria to identify suspicious reviews to be false.

Fraud detection and evil / fraudulent comments is analyzed from the point of view of the customer, the hotel or the technical side.

The purchase decision is increasingly influenced by the online consumer reviews (Ott et al., 2012). Therefore, there is growing concern about the possibility of finding false comments (deceptive opinion spam), that is, false opinions deliberately written to look authentic with the intent to deceive the reader (Ott, Choi, Cardie, & Hancock, 2011)

Ott et al. (2012), in spite of these practices having received considerable attention and concern little is known about the actual prevalence or number of false website reviews, and even less on the factors that motivate them. These authors use a model to explore the prevalence of false positive reviews (not considering the possibility of false negative feedback) in six popular communities online reviews: Expedia, Hotels.com, Orbitz, Priceline, TripAdvisor and Yelp, concluding that sites with low difficulty writing comments and make them available to a large audience, have a higher level of false comments that those websites with greater difficulties to expose a comment. Their results indicate that sites like Hotels.com have an approximate rate of 2%, TripAdvisor is 4% of fraudulent comments. The authors conclude their research with psychology linking to wonder how we lie in our communications daily, according to some studies indicating that people lie between 1 and 2 times a day (Serota et al), (DePaulo et al), and wondering in this regard if some hotels publish more fake reviews while others are more honest, or is there instead a bit of both in most hotels.

# 4 THE ROLE OF TRIPADVISOR AS CONTENT GENERATOR

The empirical analysis in this paper is made from the information obtained from TripAdvisor, so we consider it necessary to first study all the information offered by this website, also making a brief review of its history.

Founded in 2000 by Stephen Kaufer and headquartered in Newton, Massachusetts, TripAdvisor inc. listed on Nasdaq since December 2011.

With a reported income of $ 762,966,000 in 2012 and EBITDA of $ 352,474,000, 77% of its revenues come from PPC (pay per click), only 12.33% advertising (banners ...) and 10.6% from other sources (transactions, subscriptions ...)

The evolution of TripAdvisor is certainly spectacular, not only in economic and financial terms, but also in terms of visitor traffic and number of comments written.

To illustrate this growth in web traffic we can recall that in July 2006, TripAdvisor had more than 5 million reviews and opinions about hotels and attractions around 220,000. It currently receives over 200 million new visitors per month and provides users with:

• More than 100 million traveler reviews & opinions from real travelers around the world
• More than 2,500,000 business
• More than 116,000 destinations
• More than 1,100,000 lodgings
o More than 700,000 hotels
o More than 400,000 vacation rentals
• Over 259,000 attractions
• More than 24,000 restaurants in 1,100,000 cities
• 14 million traveler photos

Beyond comments, TripAdvisor websites also include links to the websites of their clients, allowing travelers to book directly with their chosen service provider (Hotels.com, Booking.com, etc..).

In addition to the flagship TripAdvisor brand has 20 other travel-related brands (smartertravel.com, travel-library.com, onetime.com, wanderfly.com ...)
A December 31, 2012 TripAdvisor had webs in 30 countries and in 21 languages and 1,575 employees of which approximately 925 are in the USA and 650 in the rest of the world.

According to Alexa.com TripAdvisor.com ranked 233 in the ranking of most viewed pages worldwide.

Although Spain is best known to those skilled in opinions on hotels and restaurants worldwide TripAdvisor´s page is more popular than yelp.com. Ranked by Alexa (traffic globally) Yelp.com ranks A level 193 since Yelp.com USA ranks 47 and Index 101

In Spain, TripAdvisor.es is much more popular than yelp, occupying the first place among the 137 most visited pages and the second since 1037.

54% of visitors are Spanish TripAdvisor.es followed by Mexican (8.5%), Peru (4.6%) and Argentina (4.5%)


Graph 1 shows the comparison in terms of traffic worldwide among TripAdvisor.com and yelp.com
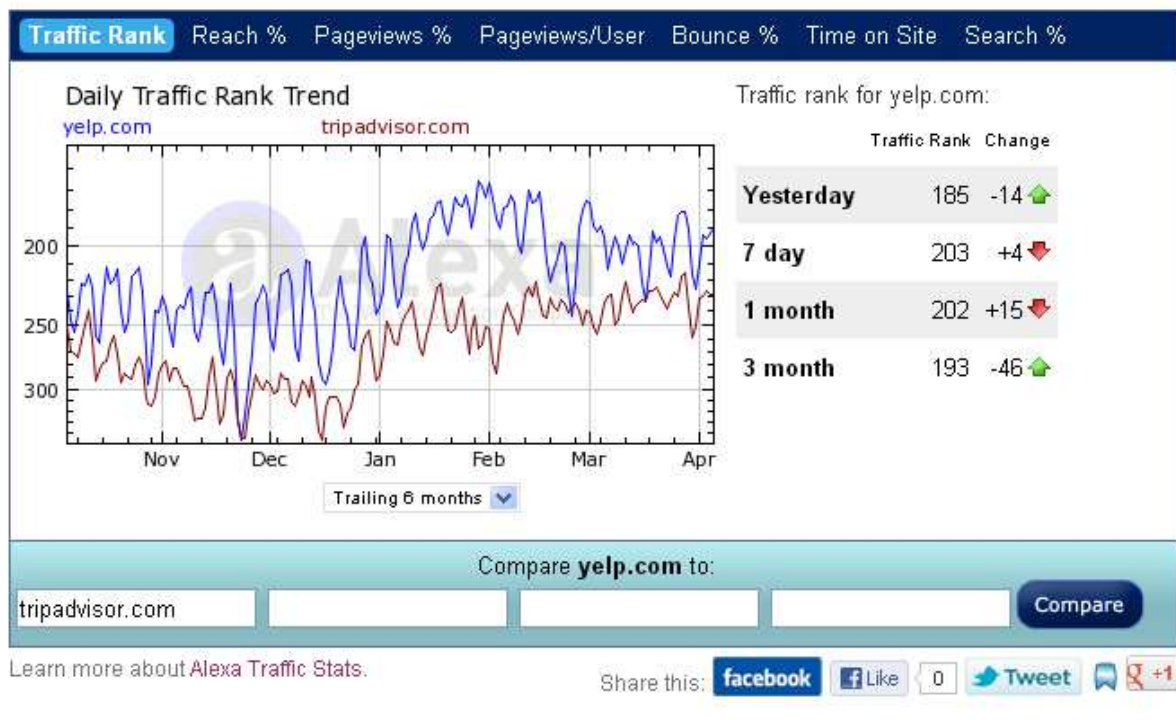


Graph 1


## 4.1. WHAT WE CAN FIND ON TRIPADVISOR?

There TripAdvisor User Review:

1- Accommodation:

• Hotels
• B & B / Inns
• Other Lodging Resorts (Apartahoteles)
• Apartments for rent

2- Restaurants

3- What to do ...

I-Attractions
• Cultural
• Museums
• Outdoors
• Actions
• Landmarks
• Sports
• Zoos and Aquariums

II-Activities
• Sightseeing
• Food and drink
• Adventures
• Wellness & Spa
• Classes

III-Nightlife

IV-Shopping

## 4.2. QUANTITATIVE AND QUALITATIVE DATA OF HOTELS

### 4.2.1 Details about hotels

The hotels can be categorized (filtered) by different criteria

1. Price range:

2. Property type:

     - Hotels,

- B & B / Inns,

- Other / apartments

3. Hotel chains

4. Stars

5. Rating (1 to 5)

For each hotel we have the following data:

1. Hotel Name

2. Location

3. Star Rating

4. Overall Media Score: Customers can give an individual hotel the following scores: Terrible (1) - Poor (2) - Normal (3) - Very Good (4) - Excellent (5). However, according to the Popularity Index of Hotels, a number between 1 and 5 is assigned to each facility, determined by the date (the later the comment is made, the greater its weight), quality and quantity of the comments and opinions recorded in TripAdvisor.

5. The position it occupies in the ranking: This ranking classifies establishments in a locality (eg hotels in Málaga) based on the mean score. This position is dynamic and varies over time. A property may have several positions depending on the geographical area of search. For example, the number one hotel in Marbella (June 2013) ranked 2nd on the Costa del Sol, 3rd in the Province of Málaga and 4th in Andalusia. (No ranking nationally).

6. Number of photos uploaded by visitors: Customers can post their pictures to highlight the benefits or problems and defects found in the hotel (dirty room, shower shabby ...)

7. Total number of hotel reviews written by users. This figure includes both positive and negative reviews.

In addition, the total number of reviews can be segmented into categories:

By level of review:

• Number of reviews as "Excellent" (punctuated with 5)

• Number of reviews as "Very Good" (scored a 4)

• Number of reviews as "Normal" (scored a 3)

• Number of reviews as "Bad" (scored with 2)

• Number of reviews as "Terrible" (scored with 1)

By type of tourist:

• Number of opinions by "Family"

• Number of opinions by "Couples"

• Number of opinions by "Solo"

• Number of opinions by "Business"

• Number of opinions by "With Friends"

8. The price of the room TripAdvisor presented for each hotel, the minimum price and the maximum among all the available rates. Recently (May 2013) the way this information is presented for each query and now a hotel and date of stay by a metasearch engine shows the lowest price available at major OTA's (Expedia.es, Hoteles.com, Booking com ...)

## 4.2.2 Data on Client

Public data about the publisher are:

1. Name (Username). Sometimes listed as anonymous

2. Hometown

3. Total Number of opinions made on TripAdvisor, which are classified as:

     - Number of opinions about hotel

- Number of reviews in cities

4. Number of useful votes received on their reviews

5. Category (Critical, Senior Critic, Senior Partner. Reviewer ...)

6.  Age Range

7.  Gender

8.  Profile of traveler. To better meet the tastes of people, we can find information on:

     - My travel style: Splurge, Splurge once in when ...

     - When I travel: Holidays, weekends ...

     - I usually travel for: Fun, work ...

     - A great holiday includes: Outdoor / Adventure, Great food / Wine ...

     - I travel with: Family with children, spouse, business, pets ...

### 4.2.3 Data on the opinions

Each written opinion includes the following information:

1. Title. A phrase or a word that sums up the customer comment. From a "Great hotel" or "Highly Recommended" to a "passable for one night", "Just enough" or even a definitive "Dismal" ...

2. Comment. Unlike other sites, such as as Booking.com, where there is a space for positive comments and a different one for negative ones, TripAdvisor makes a single field available to write the comment and sign it. In many cases the comments are hybrids, mixing in the wording of the text along with other positive and negative information. Comments are of varied length from a brief phrase to several detailed paragraphs.

3. Date and type of comment

4. Date the client stayed at the establishment.

5. Overall score: The first thing for a customer to do is to give a general view of the property using the following scale:

    1 = Terrible
    2 = Bad
    3 = Normal
    4 = Very Good
    5 = Excellent

6. Specific Ratings. You can also independently give prior valuation, specific scores on specific services (always using the same overall rating scale of 1 to 5):

    - Score on the value
    - Location Rating
    - Score on sleep quality
    - Room Rating
    - Score on cleaning
    - Score on the service
    - Score on the spa

7. Response to commentary about the hotel. Although it is unusual that hotels respond to comments, more often we find both positive and negative feedback receives a reply from the hotel management.

8. Person responsible for writing the hotel`s reply.

9. Date on which the customer`s comment was answered.

10. Tip rooms. You can make some more specific comments on a particular room (identifying number or name) for any particular feature (such as good views) to help travelers to choose a good room.

11.You can include a photo

## 4.3. TRIPADVISOR AND FRAUD

One of the key factors for the success of TripAdvisor is credibility, understanding in relation to the veracity of the events described but also in relation to the identity of the writer of the comment.

However, one of the most frequent criticisms of TripAdvisor is their lack of credibility. Although the company takes great effort to eliminate such communication and more or less widespread perception in the sector, it is true that unlike other portals like Booking.com where only users who had previously booked a room through this website were allowed to post comments, TripAdvisor for the moment allows anyone to write a review, without having to prove that they actually stayed at the hotel being commented on.

This lack of controls over entry of comments undermines the credibility of the site as it facilitates the publication of fraudulent comments of all kinds, both positive to try to improve the hotel's position in the ranking, and negative to damage the reputation for any reason (competitors, employees, customers ...)

There have been cases of customers who blackmail to threaten to write a bad review if you do not meet their expectations, as well as hotels that have been sanctioned to discover who wrote fake positive reviews to raise the hotel`s rankings.

The existence of companies that write false positive reviews in exchange for money is common knowledge in the industry.

TripAdvisor, a priori, simply rely on the good faith of the writer, with a simple check box that uses text mining techniques to detect fraud. The information provided by cookies or IP also allows you to detect fraud. They can also contact the customer or the hotel when complaining that the comment did not reflect reality and therefore fraud is suspected. Finally, the number of comments encouraging customers to participate is the best source of credibility and the best tool to dilute the influence of those fraudulent overly positive or negative comments.

## 4.4. TRIPADVISOR AND TOURISM DESTINATION

The Tourist destination, with capital as a key protagonist of the planned experience, lived and remembered, is not considered by TripAdvisor as a "valuable" subject. While it is true that we can review the attractions (La Alcazaba of Málaga, Plaza of Spain in Seville, etc.), as tourists can not give a rating to Malaga, Costa del Sol, Andalusia or Spain. Up to four tourism brands can be in our country related to a city, all with their corresponding public bodies for the promotion and creation of own brand image.

Therefore it is pertinent to know the perception that tourists have of each of these brands. And while these agencies resort to various studies and analyses to determine the positioning of its tourism brand, the fact remains that there are still few cases of agencies responsible for tourism promotion using online reputation tools for what is said in forums, social networks and their destiny beyond your website, your Twitter profile or Facebook.

But TripAdvisor is aware of the leading role of fate in the decision making process of tourists, it uses it as a communication tool to launch media campaigns or attract people to its own page, very effective in capturing the attention of possible tourists. Destinations are based on campaigns such as "Top 10 destinations for sun and sand" or "Top 10 cities to visit", but we can not assess them or know the criteria used to establish these rankings.

## 4.5. TRIPADVISOR AS A SOURCE OF INFORMATION

The comments and opinions of TripAdvisor tourists have a strong quantitative component (star position in the ranking, ratings, number of reviews ...) but they also have an equally important qualitative information load. Analysis of texts, for example, enables us to discover those issues of most concern to customers (cleanliness, location, customer ...). These comments present to professionals a source of invaluable information to learn what guests are saying. So, they could see how an operational decision such as reducing staff in the restaurant or on the counter affects the customer experience. If after that decision negative comments complaining about slow service or queues when you want to pay appear, the management could react quickly to correct the problem.

Thanks to Web 2.0, we have gone from having little information on the opinions of customers, to having an almost unimaginable amount of information, meaning if they do not have adequate tools such information becomes incomprehensible. Processing the information on the experiences and opinions of customers correctly will allow companies to improve production processes, which is a great asset for companies that know that having all this information that customers get for free at your fingertips is profitable. Companies that understand the importance of Big Data and Data Mining in the new scenario 2.0 gain invaluable competitive advantages.
One example, NH Hoteles, compiled in 2011 a total of 250,000 comments and customer feedback gathered through surveys and through the top 11 pages of comments on the Internet (TripAdvisor, Booking, Review, Zoover ...). From this information you can know what the customers` experience in areas such as breakfast or cleaning was, getting your own rankings for hotels in the chain. Information was also obtained to compare with the competition discovering what values and what experience "the hotel opposite" provides, to help them improve.

This requires recourse to sentiment analysis to extract the opinion beyond a single number on the experience of the customer through their texts. TripAdvisor Allows customers to post their own photographs to illustrate an aspect of the hotel (positive or negative), the images can be analyzed to know the most important aspects of a property.

## 4.6 QUESTIONS AND SIGNS

After analyzing hotel lists in multiple locations, small establishments (it might be more accurate to say very small) and almost always independent (non-major brands), with few stars that frequently appear at the top call our attention.

Take some examples to illustrate our assumptions (see Table 1):

• The hotel is ranked No. 1 out of 516 on the Costa del Sol, it is a property with 8 rooms and 3 stars (Hotel Los Castanos Cartajima). No hotels with over 35 rooms in the top 10 ranking Index for Costa del Sol, and the average size of the 10 rooms 13.6. The first hotel with over 100 rooms is

the Marriott's Marbella Beach Resort (288 rooms) in 16th place.

• The Hotel occupies the No. 1 place in Barcelona (Casa Camper) is a 4 star hotel with 25 rooms. The Arts Hotel (five-star with 483 rooms) is ranked No. 17.

• The No. 1 in Marbella (Hotel La Villa Marbella) is a 3 star 25 rooms. The No. 2 (The Marbella Heights Boutique Hotel), 5 rooms. The No. 6 (La Morada Mas Hermosa Hotel), 2 stars and 7 rooms. This hotel is located in the list right after the legendary Marbella Club (No. 5) and ahead of hotels as Vincci Selección Estrella del Mar (No. 8) five star hotel with 137 rooms and the Hotel Puente Romano (No. 11) Five Star and 285 rooms

Table 1: Top 10 places in the ranking Tripadvisor
Costa del Sol, Barcelona and Marbella

| Location | Ranking | Rooms | Stars | Name |
|---|---|---|---|---|
| Costa del Sol | 1 | 8 | 3 | Hotel Los Castaños (Cartajima) |
| Costa del Sol | 2 | 25 | n.d. | La Villa Marbella Charming Hotel (Marbella) |
| Costa del Sol | 3 | 9 | 3 | Hotel La Casa (Torrox) |
| Costa del Sol | 4 | 14 | 5 | Hotel Casa Rosa (Benalmádena) |
| Costa del Sol | 5 | 9 | 4 | Hotel La Luna Blanca (Torremolinos) |
| Costa del Sol | 6 | 35 | 2 | Vincci Selección Aleysa (Benalmádena) |
| Costa del Sol | 7 | 5 | 3 | The Marbella Heights Boutique Hotel (Marbella) |
| Costa del Sol | 8 | 15 | 5 | La Posada del Angel (Ojén) |
| Costa del Sol | 9 | 7 | 4 | Hotel Claude (Marbella) |
| Costa del Sol | 10 | 9 | 4 | Hotel Palacio Blanco (Vélez-Málaga) |
| Average Top 3 Costa del Sol | | 14,00 | 3,00 | |
| Average Top 5 Costa del Sol | | 13,00 | 3,75 | |
| Average Top 10 Costa del Sol | | 13,60 | 3,67 | |
| Barcelona | 1 | 25 | 4 | Csa Camper Hotel Barcelona |
| Barcelona | 2 | 72 | 5 | Alma Barcelona |
| Barcelona | 3 | 28 | 5 | Mercer Hotel Barcelona |
| Barcelona | 4 | 30 | 4 | Hotel Primero Primera |
| Barcelona | 5 | 15 | 5 | ABaC Barcelona |
| Barcelona | 6 | 92 | 3 | K+K Hotel Picasso |
| Barcelona | 7 | 169 | 4 | Hotel 1898 |
| Barcelona | 8 | 18 | 5 | Hotel DO Placa Reial |
| Barcelona | 9 | 50 | 4 | Hotel Montecarlo Barcelona |
| Barcelona | 10 | 125 | 5 | El Palace Hotel |
| Average Top 3 Barcelona | | 41,67 | 4,67 | |
| Average Top 5 Barcelona | | 34,00 | 4,60 | |
| Average Top 10 Barcelona | | 62,40 | 4,40 | |
| Marbella | 1 | 25 | 3 | La Villa Marbella Charming Hotel |
| Marbella | 2 | 5 | n.d. | The Marbella Heights Boutique Hotel |
| Marbella | 3 | 7 | 3 | Hotel Claude Marbella |
| Marbella | 4 | 288 | 5 | Marriott's Marbella Beach Resort |

| Marbella | 5 | 121 | 4 | Marbella Club Hotel |
|---|---|---|---|---|
| Marbella | 6 | 7 | 2 | La Morada Mas Hermosa Hotel |
| Marbella | 7 | 9 | 3 | Hotel Apart. Puerta de Aduares |
| Marbella | 8 | 137 | 5 | Vincci Selección Estrella del Mar |
| Marbella | 9 | 170 | 4 | Iberostar Marbella Coral Beach |
| Marbella | 10 | 180 | 4 | Hotel Guadalmina Spa & Golf Resort |
| Average Top 3 Marbella | | 12,33 | 3,00 | |
| Average Top 5 Marbella | | 89,20 | 3,75 | |
| Average Top 10 Marbella | | 94,90 | 3,67 | |

Source: Compiled from data extracted from www.TripAdvisor.es (June, -2013)

These are striking examples, we should not view them as merely exceptional cases, it seems more than justified to ask new questions:

• How is it possible to explain that on the Costa del Sol, the # 1 hotel is a 3 star hotel in Cartajima with 8 rooms, or the top 10 hotels have an average size of 13.6 rooms, and an average of 3.67 stars?

• How is it possible to explain that the average size of the top 5 hotels in Barcelona is 35 rooms?
• How does the TripAdvisor algorithm for establishments like La Villa Marbella, a Charming 3 star hotel with 25 rooms or the Hotel Claude Marbella, a 3 star hotel with 7 rooms are several places above hotels such as the Marbella Club, Puente Romano or Vincci selection Estrella del Mar?

• What is a really useful tool for decision-making is ranking in which iconic hotels such as the Arts of Barcelona or Don Pepe in Marbella are behind two and three-star hotels, or behind hotels with five to seven rooms?

• Or maybe we are not aware that many of these mythical hotel establishments have lost their levels of quality, ability to love customers and have become mediocre and predictable hotels?


## 4.7 WHAT IS TRIPADVISOR MEASURING?


A five-room hotel, just needs to provide perfect service to five clients a day in order to be exquisite. It therefore seems disproportionate to compare comments and ratings of customers of a hotel with 200 rooms with one with 7.

Weighing the same as the views of the guests of Hotel El Fuerte Marbella, No. 14 on the list, with four stars, 263 rooms, the Hotel Claude, No. 3, and 7 three star rooms can distort reality. The fact that the first has 234 views as excellent, 161 as very good, 46 as normal, 21 as bad, as awful and 14, while the second has 59 comments as excellent, 3 as very good, 2 as normal, 1 as bad and 3 being poor, can confirm our predictions about the TripAdvisor algorithm as 234 excellent opinions for hotel El Fuerte represent only 49% of all opinions while for the Hotel Claude 59 comments as excellent account for 87 % of comments.

We can guess that a small hotel, with fewer rooms, less customers, less comments, can achieve a higher percentage of excellent reviews and consequently a better position in the list than a big

hotel with many rooms, many clients and many comments .

In fact many of the reviews on these small establishments mention the treatment received directly from owner. An industry professional told us that one of these hotels (with less than 10 rooms) located in the ranking among the top 10 in Marbella, asks his clients in the booking form for their favorite song, so that upon arrival the customer is greeted with that music. This is impossible in a 200-room hotel.

How has this tool changed the way of doing things in tourism? Perhaps small hotels are favoured compared to large hotels? Try to find out.


# 5 QUANTITATIVE ANALYSIS METHODS

Despite such interesting possibilities of qualitative research in the field of online comments, our work will focus exclusively on quantitative analysis.


## 5.1. METHODOLOGICAL JUSTIFICATION

To carry out our empirical analysis we will extract data from TripAdvisor.es on all available hotels in destinations along coast of southern Spain and Portugal through an automated web crawler application.

Countless authors have used the available pages of comments, opinions and assessments of consumers and users to do their research, studying a wide range of topics and sectors from


Researching into tourism issues, students increasingly find more publications about local areas in Oceania, Asia or Europe, most studies that use data from TripAdvisor, Booking or Expedia are located in USA, such as Jurca, (2010) International data extracted from all hotels in 4 cities (Boston, Las Vegas, New York and Sydney) with more than 10 comments. O'connor (2010) randomly selected 100 from the list of hotels listed on TripAdvisor 1,042 from the City of London. Cunningham et al. (2010) analyzed 30,000 comments on Irish hotels and 50,000 hotels of Las Vegas. Stringam et al (2010) extracted Expedia.com comments on all the hotels in the 100 largest cities in the USA, 60,648 comments were collected on 10,537 hotels. Finally, Callarisa et al. (2012) pulled information from 45,914 TripAdvisor reviews on Paris hotels and 8,736 on Hong Kong hotels.

Given the extensive literature on tourism research using these sources of information, we believe this amply justified from a methodological point of view TripAdvisor`s extraction of data to analyze the tourism sector.

Another important aspect is the use of IT to automate the extraction of all these data applications. Jurca, (2010) extracted the information with a web crawler. Stringam et al (2010) used an "automated web spider" to remove comments from Expedia.com and Callarisa et al. (2012) implemented an algorithm in JAVA  to extract information from TripAdvisor. Gerdes et al (2008) specifically designed spider - a web application - to extract data from the Internet.

While it is true that on occasion legal and ethical questions have been raised (Johnson et al, 2012), we understand that the literature that uses these tools to gather information is so broad that it can also be considered an acceptable technique for collecting data .

Ultimately we believe that the use of TripAdvisor as a provider of information and automated

system for extracting it are two characteristics of our work that are both widely used and accepted in the scientific literature.

## *5.2 DESIGN OF DATABASE*

We have collected the following information for each of the hotels located in coastal towns of southern Spain and Portugal:

Data for each Hotel:

1 Hotel Name
2 Address
3 Post Code
4 Tourist Point: Algarve, Costa del Sol, Costa Brava ...
5 Tourist Zone: Málaga, Marbella, Salou, ...
6 Tourist spot ranking position
7 Ranking Position in the Tourist Zone
8 Overall Media Score
9 Number of reviews
10 Stars
11 Number of Rooms
12 Price
13 Number of Reviews "Excellent" (5)
14 Number of reviews "Very Good" (4)
15 Number of reviews "Normal" (3)
16 Number of reviews "Bad" (2)
17 Number of reviews "Terrible" (1)
18 Number of opinions in "Family" category
19 Number of opinions in "Romance" category
20 Number of opinions in "Solitude" category
21 Number of opinions in "Business" category
22 Number of opinions in "With Friends" category
23 Sleep quality
24 Value for money
25 Services
26 Rooms
27 Location
28 Cleaning

From this data, collected directly and automatically, a series of calculated fields were created to add to the database:

29 Percentage of opinions "Excellent" out of all reviews
30 Percentage of opinions "Very Good" out of all reviews
31 Percentage of opinions "Normal" out of all reviews
32 Percentage of opinions "Bad" out of all reviews
33 Percentage of opinions "Terrible" out of all reviews
34 Sum of Percentage of opinions "Excellent" Increased percentage of opinions "Very Good" and "Outstanding"
35 Sum of Percentage of opinions "Bad" Increased percentage of opinions "Terrible" (Poor)
36 Difference between the percentage of "Outstanding" opinions minus the percentage of "Poor" Reviews. This value will vary from -1 to +1
37 Percentage of opinions "En Familia" on total reviews

38 Percentage of opinions "In Romance" on total reviews
39 Percentage of opinions "In Solitude" on total reviews
40 Percentage of opinions "Business" of total reviews
41 Percentage of opinions "With Friends" on total reviews
42 Number of hotel reviews divided by the number of rooms. Thus we get the average number of comments written per room.
43 Number of comments "Excellent + Very good" per hotel divided by the number of rooms. There were as many "good" comments written on average per room.
44 Number of comments "Malo + Terrible" per hotel divided by the number of rooms. Thus we get a few "bad" reviews written on average per room

Data has been compiled for all hotels in the following destinations

Table 2: Number of Initial Establishments

| Order | Location | Initial Hotels in our sample | % Sample |
|---|---|---|---|
| 1 | Islas Baleares | 556 | 17,36% |
| 2 | Costa de Barcelona | 537 | 16,77% |
| 3 | Islas Canarias | 380 | 11,87% |
| 4 | Algarve | 281 | 8,78% |
| 5 | Costa del Sol | 281 | 8,78% |
| 6 | Costa Blanca | 258 | 8,06% |
| 7 | Costa Brava | 216 | 6,75% |
| 8 | Costa Dorada | 153 | 4,78% |
| 9 | Costa de la Luz | 114 | 3,56% |
| 10 | Costa de Valencia | 109 | 3,40% |
| 11 | Costa Almería | 100 | 3,12% |
| 12 | Costa de Murcia | 100 | 3,12% |
| 13 | Costa de Castellón | 90 | 2,81% |
| 14 | Costa Tropical | 27 | 0,84% |
| | TOTAL | 3.202 | 100,00% |

Initially we obtained information from 3,202 hotels, but we have removed the database for those establishments that did not have the figure for the total number of views, ie to eliminate 568 hotels (17.7% of the records). We have also eliminated those establishments with less than 20 comments, and we understand that if our work will especially focus on the quantitative analysis of these comments, the hotels with such small numbers can skew results.

This procedure is supported by the scientific literature when multiple authors limit their analysis to hotels with a minimum of comments. For example (Silveira et al.2012) selected a sample of small hotels in Portugal (less than 120 rooms) with at least 30 reviews while Marchiori et al, (2011) selected 77 Swiss hotels with at least 200 comments.

In our case we have eliminated hotels with less than 30 comments which has meant giving up 598 hotels (the 18.67% of the sample), so we decided to lower the number to 20 comments which meant eliminating 423 hotels, the 13.21% of the sample.

Finally we removed the data presented in the following table:

Table 3: Data Deleted

|  | Deleted Hotels | % Sample |
|---|---|---|
| Hotels with no reviews | 568 | 17,74% |
| < 20 reviews | 423 | 13,21% |
| Total | 991 | 30,95% |

Our final database is therefore composed of 2,211 hotels.
INE terminology is used to distinguish between Tourism Zones and Tourist Spots and the database can be grouped by any of these two terms.

So can Tourist Zones shown in the table below:

Table 4: Hotels Tourist Zones

| Order | Location (Tourist zones) | Hotels (with 20 or more reviews) in our sample | Sample % |
|---|---|---|---|
| 1 | Costa de Barcelona | 413 | 18,68% |
| 2 | Islas Baleares | 385 | 17,41% |
| 3 | Islas Canarias | 337 | 15,24% |
| 4 | Costa del Sol | 233 | 10,54% |
| 5 | Costa Blanca | 187 | 8,46% |
| 6 | Algarve | 147 | 6,65% |
| 7 | Costa Brava | 138 | 6,24% |
| 8 | Costa Dorada | 123 | 5,56% |
| 9 | Costa de la Luz | 80 | 3,62% |
| 10 | Costa de Murcia | 61 | 2,76% |
| 11 | Costa de Valencia | 36 | 1,63% |
| 12 | Costa de Castellón | 31 | 1,40% |
| 13 | Costa Almería | 22 | 1,00% |
| 14 | Costa Tropical | 18 | 0,81% |
|  | TOTAL | 2.211 | 100,00% |

Data on Tourist Spots is presented in the following table:

Table 5: Hotels in Tourist Spots

| Order | Location (Tourist Spots) | Hotels(with 20 or more reviews) | % Sample |
|---|---|---|---|
| 1 | Barcelona | 328 | 14,83% |
| 2 | Benidorm | 119 | 5,38% |
| 3 | Calvià | 83 | 3,75% |

| 4 | Lloret de Mar | 70 | 3,17% |
|---|---|---|---|
| 5 | Playa del Inglés Gran Canaria | 63 | 2,85% |
| 6 | Sant Antoni de Portmany | 63 | 2,85% |
| 7 | Salou | 61 | 2,76% |
| 8 | Puerto del Carmen Lanzarote | 54 | 2,44% |
| 9 | Puerto de la Cruz Tenerife | 53 | 2,40% |
| 10 | Albufeira | 52 | 2,35% |
| 11 | Torremolinos | 51 | 2,31% |
| 12 | Marbella | 48 | 2,17% |
| 13 | Málaga | 48 | 2,17% |
| 14 | Playa de Palma | 47 | 2,13% |
| 15 | Palma de Mallorca | 45 | 2,04% |
| 16 | Jandía Fuerteventura | 42 | 1,90% |
| 17 | Playa de las Américas Tenerife | 38 | 1,72% |
| 18 | Maspalomas Gran Canaria | 36 | 1,63% |
| 19 | Benalmádena | 34 | 1,54% |
| 20 | Cala Millor | 32 | 1,45% |
| 21 | S'Arenal | 31 | 1,40% |
| 22 | Port d'Alcudia | 30 | 1,36% |
| 23 | Alicante | 30 | 1,36% |
| 24 | Cala d'Or | 28 | 1,27% |
| 25 | Tossa de Mar | 28 | 1,27% |
| 26 | Calella | 27 | 1,22% |
| 27 | Corralejo Fuerteventura | 26 | 1,18% |
| 28 | Sitges | 26 | 1,18% |
| 29 | Ciutadella | 25 | 1,13% |
| 30 | Puerto Rico Gran Canaria | 25 | 1,13% |
| 31 | Valencia | 23 | 1,04% |
| 32 | Vilamoura | 22 | 1,00% |
| 33 | Roses | 21 | 0,95% |
| 34 | Fuengirola | 20 | 0,90% |
| 35 | Nerja | 19 | 0,86% |
| 36 | Peñíscola | 17 | 0,77% |
| 37 | Santa Susanna | 17 | 0,77% |
| 38 | Denia | 17 | 0,77% |
| 39 | Cambrils | 16 | 0,72% |
| 40 | Malgrat de Mar | 15 | 0,68% |
| 41 | Murcia | 15 | 0,68% |
| 42 | Tarragona | 15 | 0,68% |
| 43 | Tarifa | 14 | 0,63% |
| 44 | Almuñecar | 13 | 0,59% |
| 45 | Praia da rocha | 13 | 0,59% |
| 46 | Cádiz | 13 | 0,59% |
| 47 | El Puerto de Santa María | 13 | 0,59% |
| 48 | Faro | 13 | 0,59% |
| 49 | Conil | 12 | 0,54% |
| 50 | Blanes | 11 | 0,50% |
| 51 | La Manga del Mar Menor | 11 | 0,50% |
| 52 | Portimao | 11 | 0,50% |
| 53 | Carvoeiro | 11 | 0,50% |
| 54 | Gandia | 10 | 0,45% |
| 55 | Alvor | 10 | 0,45% |
| 56 | Mijas | 10 | 0,45% |
| 57 | Calpe | 9 | 0,41% |
| 58 | L'Estartit | 8 | 0,36% |
| 59 | Quarteira | 8 | 0,36% |

| | | | |
|---|---|---|---|
| 60 | Zahara de los atunes | 8 | 0,36% |
| 61 | Ayamonte | 7 | 0,32% |
| 62 | L'Alfas del Pi | 7 | 0,32% |
| 63 | La Pineda | 7 | 0,32% |
| 64 | Almancil | 7 | 0,32% |
| 65 | Benicasim | 6 | 0,27% |
| 66 | Cartagena | 6 | 0,27% |
| 67 | Huelva | 6 | 0,27% |
| 68 | Calafell | 5 | 0,23% |
| 69 | Coma Ruga | 5 | 0,23% |
| 70 | Mojacar | 5 | 0,23% |
| 71 | San Pedro del Pinatar | 5 | 0,23% |
| 72 | Guardamar del Segura | 5 | 0,23% |
| 73 | Alcossebre | 4 | 0,18% |
| 74 | Archena | 4 | 0,18% |
| 75 | Los Alcázares | 4 | 0,18% |
| 76 | Roquetas de Mar | 4 | 0,18% |
| 77 | Almería | 3 | 0,14% |
| 78 | Cartaya | 3 | 0,14% |
| 79 | Cullera | 3 | 0,14% |
| 80 | Mazarrón | 3 | 0,14% |
| 81 | Punta Umbría | 3 | 0,14% |
| 82 | Reus | 3 | 0,14% |
| 83 | San José | 3 | 0,14% |
| 84 | Torre-Pacheco | 3 | 0,14% |
| 85 | Manilva | 3 | 0,14% |
| 86 | Aguadulce | 2 | 0,09% |
| 87 | El Toyo | 2 | 0,09% |
| 88 | L'Hospitalet de l'Infant | 2 | 0,09% |
| 89 | La Herradura | 2 | 0,09% |
| 90 | Lorca | 2 | 0,09% |
| 91 | Oropesa del Mar | 2 | 0,09% |
| 92 | Totana | 2 | 0,09% |
| 93 | Vera | 2 | 0,09% |
| 94 | Vinaròs | 2 | 0,09% |
| 95 | Águilas | 2 | 0,09% |
| 96 | Alhama de Murcia | 1 | 0,05% |
| 97 | Cala Ratjada | 1 | 0,05% |
| 98 | Calahonda | 1 | 0,05% |
| 99 | El Vendrell | 1 | 0,05% |
| 100 | L'Espluga de Francolí | 1 | 0,05% |
| 101 | La Alberca | 1 | 0,05% |
| 102 | La Antilla | 1 | 0,05% |
| 103 | La Selva del Camp | 1 | 0,05% |
| 104 | Los Belones | 1 | 0,05% |
| 105 | Los Escullos | 1 | 0,05% |
| 106 | Miami Platja | 1 | 0,05% |
| 107 | Montbrio del Camp | 1 | 0,05% |
| 108 | Motril | 1 | 0,05% |
| 109 | Pratdip | 1 | 0,05% |
| 110 | Salobrena | 1 | 0,05% |
| 111 | San Javier | 1 | 0,05% |
| 112 | Sant Salvador | 1 | 0,05% |
| 113 | Valls | 1 | 0,05% |
| 114 | Vila Seca | 1 | 0,05% |
| Total | | 2.211 | 100,00% |

## 5.3 QUANTITATIVE TECHNIQUES

To carry out our work, we will refer primarily to a descriptive statistical analysis using line charts, bar charts, line plots, histograms and box plots.
We also analyze the data using contingency tables and their corresponding horizontal and vertical percentages.
Thirdly, will make the analyses of variance of the major variables involved in this study.
Finally and for these same variables, we will make a non-parametric analysis (Kruskal-Wallis).

# 6 RESULTS

## 6.1 JOINT ANALYSIS OF COASTAL DESTINATIONS

In this first section we will not pay attention to tourist destinations from a geographical point of view as we want to study the relationship between the following variables.

1 The average hotel rating overall.

2 Hotel size (measured by number of rooms)

3 The number of comments on each hotel (in proportion to its size)

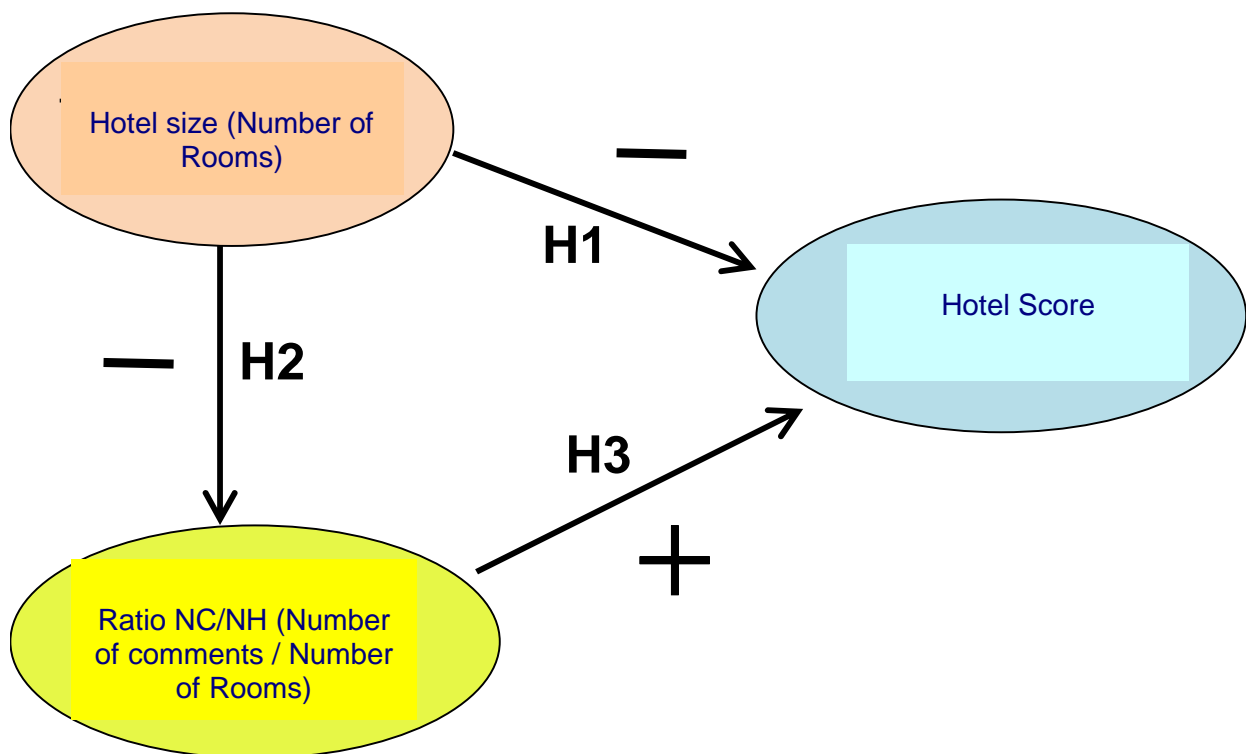In this section we answer three questions:

1 Are customers very critical when rating a hotel?

2 Is there a relationship between the size of it and score it receives from customers?

3 Is there a relationship between the number of comments received by a hotel and the mean score of the same?

Our starting hypothesis is summarized in the following graph:



Graph 2: Starting Hypothesis

H1: There is a negative relationship between the size of the hotel and the mean score of the same.
(On average small hotels have higher scores than the big hotels)

H2: There is a negative relationship between the size of the hotel and the number of comments it receives (relative to its size).
Small hotels receive on average a greater number of comments (in proportion to its size) than large hotels

H3: There is a positive relationship between the number of comments (relative to size) and the mean score of the establishment.

Hotels that achieve a higher number of comments (relative to their size) get better grades by term.

## 6.1.1 The mean overall score (PGM) of Hotels

Let's begin by performing a statistical analysis of this variable.

Customers can give an individual hotel the following scores:

Terrible (1) - Poor (2) - Normal (3) - Very Good (4) - Excellent (5)

The Global Average rating each establishment is the average of the individual rating, but is not a strictly continuous variable because the values applied to a hotel are:

1 - 1,5 - 2 - 2,5 - 3 - 3.5 - 4 - 4,5 - 5

As shown in Table 6, the mean score of the hotels in our sample is 3.72. However, the median value is even higher, reaching a 4, that is, half of the hotels have a score of 4; 4.5 or 5.

Another noteworthy fact is that while there are hotels with the highest possible score (5), there are none with a lowest (1) and the minimum value in the sample is 1.5. There are no hotels with an average overall score of 1.

Table 6: Overall Mean Score (Statistics)

|  | Value |
|---|---|
| N | 2211 |
| Mean | 3.72 |
| Standard deviation | 0,56 |
| Variance | 0,32 |
| Coef.Pearson. | 15,05% |
| Mínimum | 1,50 |
| Máximum | 5,00 |
| Median | 4,00 |

The following table shows the relative frequencies for the different observed values of the variable, with the modal value 4.

Table 7: Overall Score Media (Frequencies)

| Value | Frecuency | Percentage | Percentage Acummulated |
|---|---|---|---|
| 1.50 | 4 | 0,18% | 0,18 |
| 2.00 | 24 | 1,09% | 1,27 |
| 2.50 | 82 | 3,71% | 4,98 |
| 3.00 | 306 | 13,84% | 18,82 |
| 3.50 | 633 | 28,63% | 47,44 |
| 4.00 | 815 | 36,86% | 84,31 |
| 4.50 | 327 | 14,79% | 99,1 |
| 5.00 | 20 | 0,90% | 100 |
| Total | 2211 | 100.0 | |

The chart below shows the histogram for the different ratings
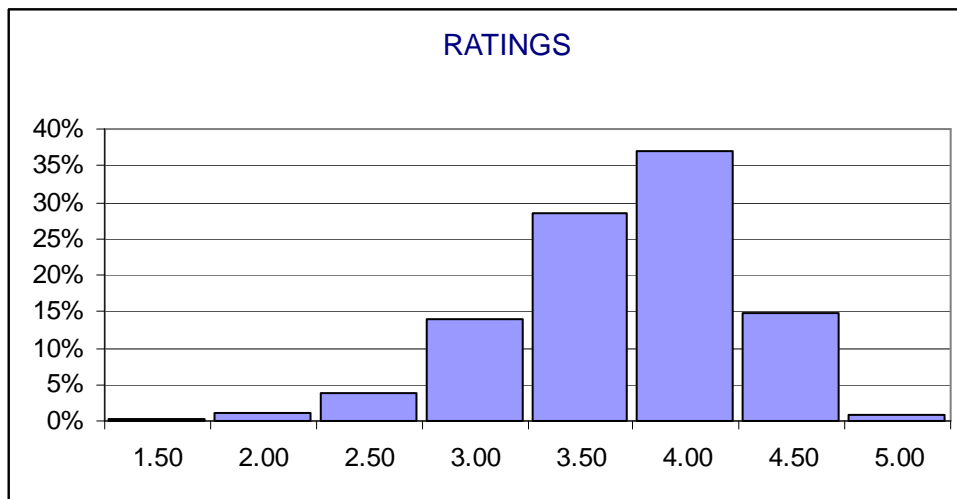
RATINGS

Graph 3

In the following table we group the scores we have called "Notables" (4 - 4,5 - 5), and we have called "Bad" (1,5 - 2 - 2,5).

Table 8: Hotels Rating High (4 - 4,5 - 5) and Low (1,5 - 2 - 2,5)

| Value | Frecuency | Percentage |
|---|---|---|
| (5 - 4,5 - 4) | 1162 | 52,56% |
| (1,5 - 2 -2,5) | 110 | 4,98% |
| (5 - 4,5) | 347 | 15,69% |
| (1,5 - 2) | 28 | 1,27% |
| (5) | 20 | 0,90% |
| (1,5) | 4 | 0,18% |
| (3 - 3,5) | 939 | 42,47% |

As we can see, the majority (52.56%) of the group of hotels have a high score (5 to 4.5 - 4) Page, while the number of hotels with low scores (1, 5-2 - 2.5) is very low (4.98%)
The percentage of hotels with "Very Good" scores (4.5 -5) is 15.69% of the total sample while the percentage of hotels with "Very Bad" scores (1.5 - 2) is just 1.27% .

Finally, both the percentage of hotels with the highest score (5) and the minimum score in the sample (1.5) are not significant (0.9% and 0.18% respectively)

The chart below shows the cumulative percentages indicating that hotels with three points or less account for only 18.8% of the sample.
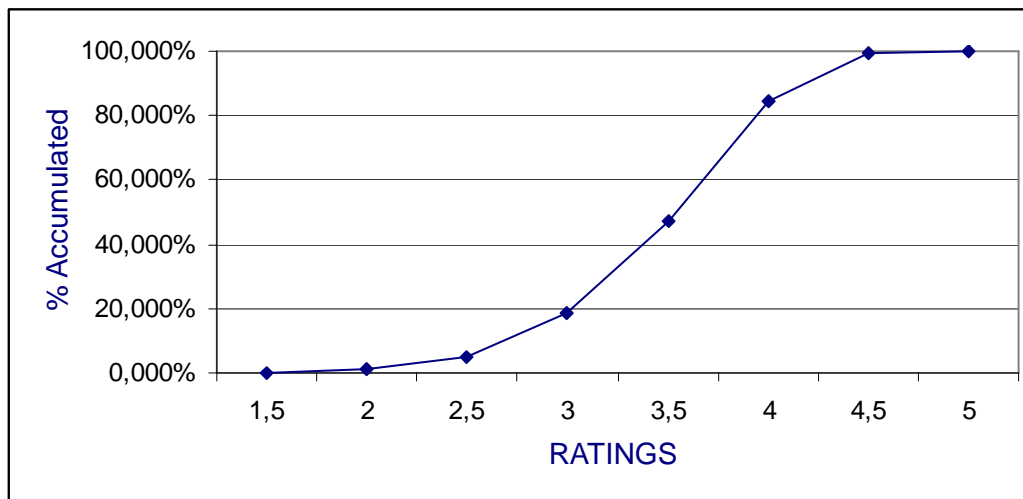
Graph 4: Ratings and cumulative frequencies

To conclude this section, we highlight as the most significant fact that most tourists favourably rate their hotel experiences.

## 6.1.2 Rooms and Ratings

The second variable we want to study in depth is the size of the hotel. We answer the following question:

Does the size of the hotel - measured by the number of rooms it has - affect the score it receives from customers?

As in the previous section, we will perform a descriptive statistical analysis of this variable.

The following table lists the main indicators

Table 9: Number of Bedrooms (Statistics)

|  | Valor |
| --- | --- |
| *N* | 2205 |
| *Average* | 149,94 |
| *Standard deviation* | 131,64 |
| *Variance* | 17328,89 |
| *Coef,Pearson* | 87,795% |
| *Mínimum* | 2 |
| *Máximum* | 1.468 |
| *Median* | 112 |

The average size of the hotels in our sample is approximately 150 rooms, while the median value is 112.

The dispersion of the variable is significant, finding hotels from 2 up to 1,468 bedrooms.

The chart below shows the frequency distribution of the Hotels Size variable.

Hotels Size (Rooms) Frequencies
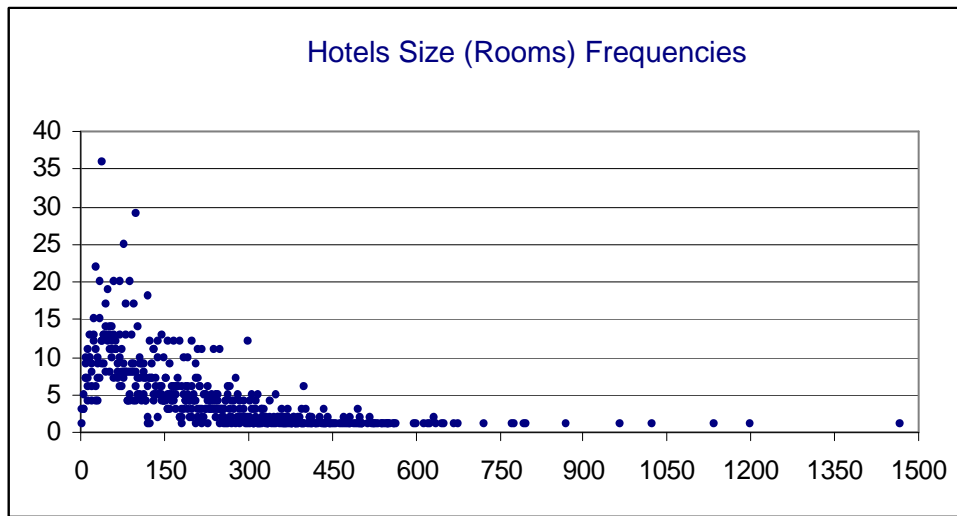
Graph 5: Type (frequencies)

Analysis of Variance

To check if there is any relationship between the number of hotel rooms and the average score on TripAdvisor, we will conduct an analysis of variance.

First homogeneous grouped observations collected at intervals of approximately 10% of the data. As shown in the following table, all observations are divided into 10 groups, the first which includes hotels that have rooms from 1 to 30 and the last those between 313 and 1,468 rooms.

The following table contains information about each class:

Table 10: Number of rooms (classes)

| Class | Value | Frequency | Average | Standard Deviation | Percentage | Percentage Accumulated |
|-------|-------|-----------|---------|--------------------|------------|------------------------|
| 1 | 1-30 | 219 | 3,9726 | 0,59175 | 9,91% | 9,93% |
| 2 | 31-48 | 228 | 3,7544 | 0,54428 | 10,31% | 20,27% |
| 3 | 49-65 | 210 | 3,7286 | 0,54037 | 9,5% | 29,8% |
| 4 | 66-88 | 226 | 3,7389 | 0,51682 | 10,22% | 40,05% |
| 5 | 89-111 | 217 | 3,6244 | 0,64060 | 9,81% | 49,89% |
| 6 | 112-144 | 224 | 3,6741 | 0,56310 | 10,13% | 60,05% |
| 7 | 145-185 | 218 | 3,7248 | 0,52357 | 9,86% | 69,93% |
| 8 | 186-233 | 224 | 3,6674 | 0,53140 | 10,13% | 80,09% |
| 9 | 234-312 | 220 | 3,6932 | 0,51258 | 9,95% | 90,07% |
| 10 | 313-1468 | 219 | 3,6164 | 0,57582 | 9,91% | 100% |
| Total | | 2205 | 100,0 | | | |

To be included in the analysis of variance three conditions must be met:

1- Class Independence
2- Normality
3- Homogeneity of variances

Then the Levene test for homogeneity of variances was conducted.

Table 11: Test of homogeneity of variances (No. rooms and Rating)

| Statistic of Levene | gl1 | gl2 | Sig. |
|---|---|---|---|
| 1,112 | 9 | 2195 | ,350 |

For Sig.Level = 0.350 > 0.05, we accept the null hypothesis (H0) of equal variances.

Then we perform the analysis of variance:

Table 12: Analysis of variance (No. rooms and Rating)

| | Sum of squares | gl | Root mean square | F | Sig. |
|---|---|---|---|---|---|
| Inter-grupos | 19,925 | 9 | 2,214 | 7,185 | ,000 |
| Intra-grupos | 676,336 | 2195 | ,308 | | |
| Total | 696,261 | 2204 | | | |

With a Sig.Level <0.05 reject H0 (equality of averages). Therefore some averages are significantly different, which indicates that there is some relationship between the number of hotel rooms and the mean score of the same.

Nonparametric test

However, and to complete the analysis of variance considering it has not under-gone the test of normality, we apply a nonparametric test (Kruskal-Wallis) to confirm the results:

Table 13: Kruskal-Wallis (Number of rooms and Rating)

| | Rating |
|---|---|
| Chi-square | 55,908 |
| Gl | 9 |
| Sig. asintót. | ,000 |
| **Statistic of contrast[a,b]** | |
| a. Kruskal-Wallis Test | |
| b. Grouping variable: clases of rooms | |

The Kruskal-Wallis test results are accurate to 5%.(Discarding H0)
This reinforces the conclusion that there are significant differences between the median levels.

Analysis de tables de contingence

Using the following tables, we analyse the horizontal and vertical percentages to understand the relationship between the number of rooms per hotel and their average global score. The following table collects the frequencies observed for each combination of possible cases.

Table 14: Table of contingence Nº of Rooms and PGM (frequencies)

| Number of Rooms (Classes) | Rating | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | 1,50 | 2,00 | 2,50 | 3,00 | 3,50 | 4,00 | 4,50 | 5,00 | |
| 1-30 | 1 | 1 | 2 | 20 | 47 | 73 | 60 | 15 | 219 |
| 31-48 | 0 | 3 | 7 | 25 | 70 | 83 | 39 | 1 | 228 |
| 49-65 | 0 | 2 | 6 | 33 | 54 | 84 | 30 | 1 | 210 |
| 66-88 | 0 | 0 | 6 | 34 | 71 | 77 | 37 | 1 | 226 |
| 89-111 | 2 | 6 | 15 | 25 | 64 | 76 | 28 | 1 | 217 |
| 112-144 | 0 | 4 | 10 | 33 | 63 | 85 | 29 | 0 | 224 |
| 145-185 | 0 | 0 | 10 | 29 | 64 | 84 | 30 | 1 | 218 |
| 186-233 | 1 | 10 | 6 | 40 | 69 | 80 | 27 | 0 | 224 |
| 234-312 | 0 | 20 | 8 | 27 | 75 | 82 | 26 | 0 | 220 |
| 313-1468 | 0 | 50 | 12 | 39 | 54 | 89 | 20 | 0 | 219 |
| Total | 4 | 24 | 82 | 305 | 631 | 813 | 326 | 20 | 2205 |

The following table shows the percentages (by row) of hotels segmented by size and their respective scores:

Table 15: Contingency table Bedrooms and PGM (Horizontal %)

| Rooms | Rating | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | 1,5 | 2 | 2,5 | 3 | 3,5 | 4 | 4,5 | 5 | |
| 1-30 | 0,46% | 0,46% | 0,91% | 9,13% | 21,46% | 33,33% | 27,40% | 6,85% | 100% |
| 31-48 | 0,00% | 1,32% | 3,07% | 10,96% | 30,70% | 36,40% | 17,11% | 0,44% | 100% |
| 49-65 | 0,00% | 0,95% | 2,86% | 15,71% | 25,71% | 40,00% | 14,29% | 0,48% | 100% |
| 66-88 | 0,00% | 0,00% | 2,65% | 15,04% | 31,42% | 34,07% | 16,37% | 0,44% | 100% |
| 89-111 | 0,92% | 2,76% | 6,91% | 11,52% | 29,49% | 35,02% | 12,90% | 0,46% | 100% |
| 112-144 | 0,00% | 1,79% | 4,46% | 14,73% | 28,13% | 37,95% | 12,95% | 0,00% | 100% |
| 145-185 | 0,00% | 0,00% | 4,59% | 13,30% | 29,36% | 38,53% | 13,76% | 0,46% | 100% |
| 186-233 | 0,45% | 0,45% | 2,68% | 17,86% | 30,80% | 35,71% | 12,05% | 0,00% | 100% |
| 234-312 | 0,00% | 0,91% | 3,64% | 12,27% | 34,09% | 37,27% | 11,82% | 0,00% | 100% |
| 313-1468 | 0,00% | 2,28% | 5,48% | 17,81% | 24,66% | 40,64% | 9,13% | 0,00% | 100% |

If we now analyse the data by columns, we find that 75% of the hotels that receive the highest score are smaller hotels (hotels with 1 to 30 rooms)

Table 16: Contingency table Bedrooms and PGM (% Vertical)

| Rooms | Rating | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *1,5* | *2* | *2,5* | *3* | *3,5* | *4* | *4,5* | *5* |
| 1-30 | 25% | 4,17% | 2,44% | 6,56% | 7,45% | 8,98% | 18,40% | 75% |
| 31-48 | 0% | 12,50% | 8,54% | 8,20% | 11,09% | 10,21% | 11,96% | 5% |
| 49-65 | 0% | 8,33% | 7,32% | 10,82% | 8,56% | 10,33% | 9,20% | 5% |
| 66-88 | 0% | 0% | 7,32% | 11,15% | 11,25% | 9,47% | 11,35% | 5% |
| 89-111 | 50% | 25% | 18,29% | 8,20% | 10,14% | 9,35% | 8,59% | 5% |
| 112-144 | 0% | 16,67% | 12,20% | 10,82% | 9,98% | 10,46% | 8,90% | 0% |
| 145-185 | 0% | 0% | 12,20% | 9,51% | 10,14% | 10,33% | 9,20% | 5% |
| 186-233 | 25% | 4,17% | 7,32% | 13,11% | 10,94% | 9,84% | 8,28% | 0% |
| 234-312 | 0% | 8,33% | 9,76% | 8,85% | 11,89% | 10,09% | 7,98% | 0% |
| 313-1468 | 0% | 20,83% | 14,63% | 12,79% | 8,56% | 10,95% | 6,13% | 0% |
| Total | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |

Technical Analysis

Finally a graphical analysis illustrates the relationship between the size of the hotels (measured by number of rooms) and the mean overall score.



---
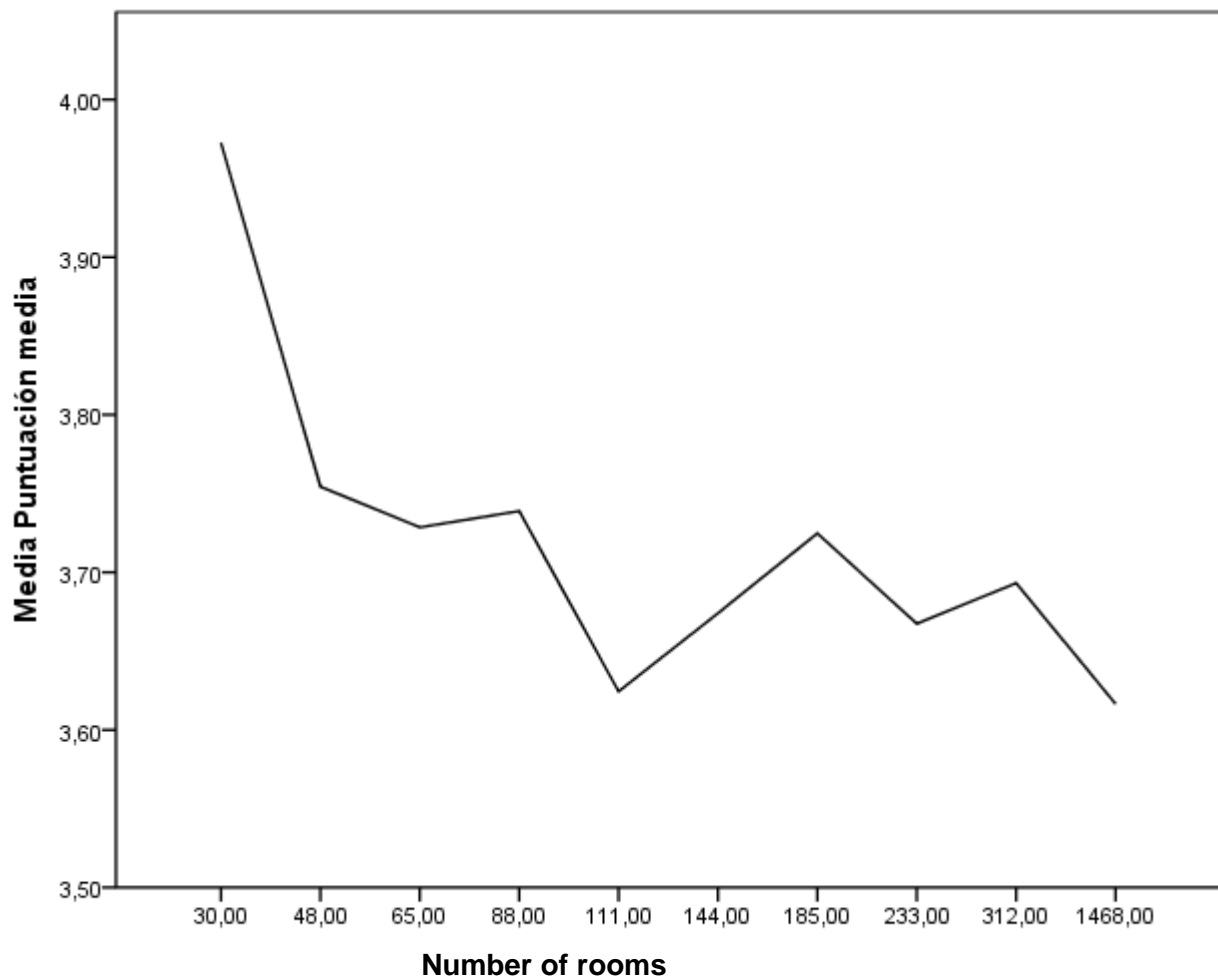
The graph shows that as a hotels` average score decreases (arithmetic mean) their size (number of rooms) increases

High Scores

Here we analyse only the highest scores.

Table 17: Overall Average Score (4 - 4,5 - 5)

| Rooms | Highest scores (4 – 4,5 – 5) Horizontal % | | | | | |
|---|---|---|---|---|---|---|
| | 4 | 4,5 | 5 | | 4,5 - 5 | 4 - 4,5 - 5 |
| 1-30 | 33,33% | 27,40% | 6,85% | | 34,25% | 67,58% |
| 31-48 | 36,40% | 17,11% | 0,44% | | 17,55% | 53,95% |
| 49-65 | 40,00% | 14,29% | 0,48% | | 14,77% | 54,77% |
| 66-88 | 34,07% | 16,37% | 0,44% | | 16,81% | 50,88% |
| 89-111 | 35,02% | 12,90% | 0,46% | | 13,36% | 48,38% |
| 112-144 | 37,95% | 12,95% | 0,00% | | 12,95% | 50,90% |
| 145-185 | 38,53% | 13,76% | 0,46% | | 14,22% | 52,75% |
| 186-233 | 35,71% | 12,05% | 0,00% | | 12,05% | 47,76% |
| 234-312 | 37,27% | 11,82% | 0,00% | | 11,82% | 49,09% |
| 313-1468 | 40,64% | 9,13% | 0,00% | | 9,13% | 49,77% |

6.85% of the younger group (1-30) of Hotels has an excellent score (5). In the rest of this group, the score ranges from less than 1%, and 0% from 186 rooms.
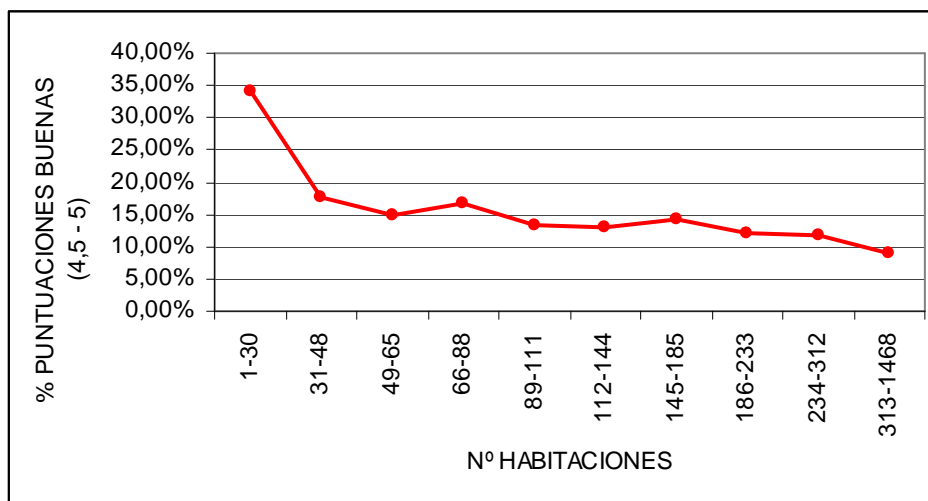34.25% of the group (1-30) hotels get a very good score (4.5 or 5), 9.13% in the group of larger hotels (313-1468).
The 67.58% of the group (1-30) hotels get a good score (4 - 4,5 or 5), as do the the 49.77% in the range of larger hotels (313-1468).
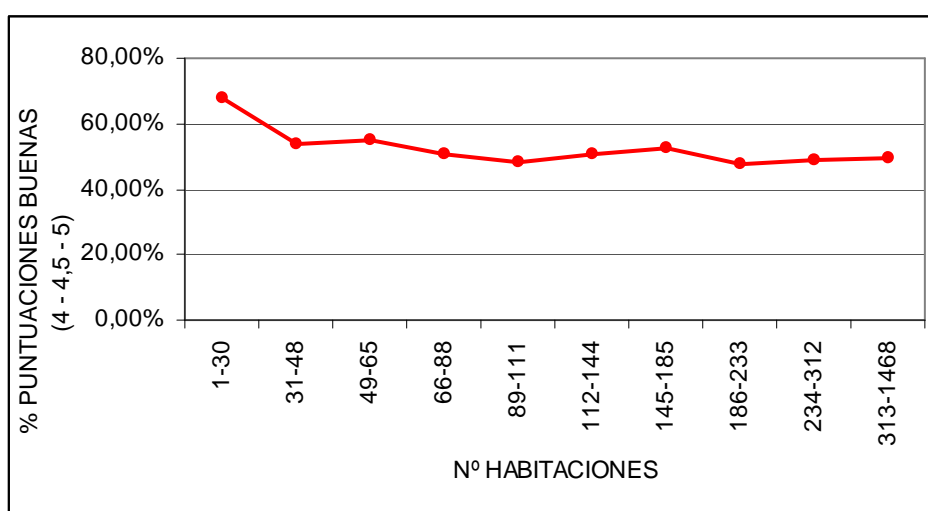
Now let's see the data graphically:
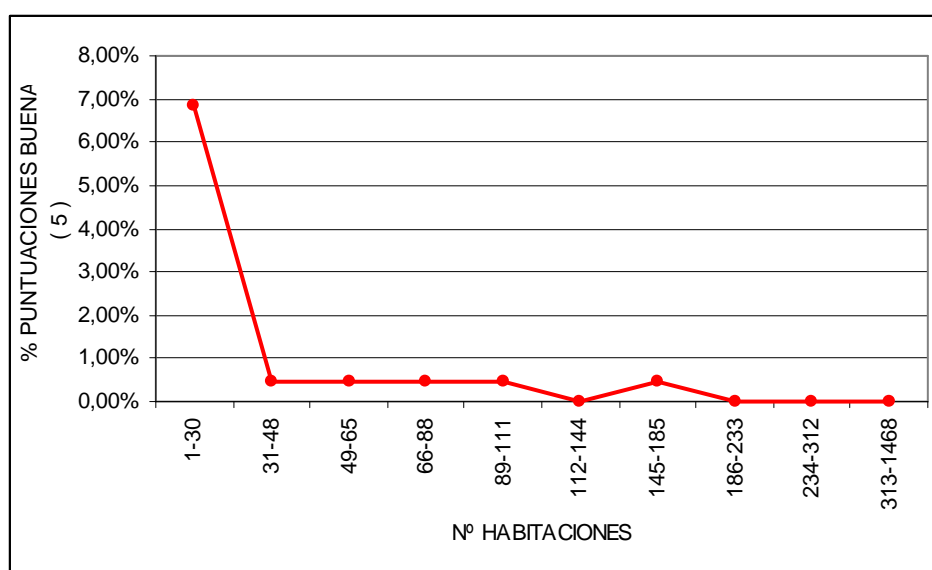


Graph 7: Highest score (5)

Graph 8: Highest scores (4.5 - 5)



Graph 9: Highest scores (4 - 4.5 - 5)

As the size of the hotel increases, there is a trend that the percentage of qualified hotels with the highest scores decreases.

Low ratings

We then performed similar analysis but for lower scores.

Table 18: Low scores (1,5 - 2 - 2.5)

|  | Scores (1,5 – 2 – 2,5) Horizontal % | | | | |
|---|---|---|---|---|---|
|  | 1,5 | 2 | 2,5 | 1,5 – 2 | 1,5 - 2 - 2,5 |
| Nº Hoteles | 4 | 24 | 82 | 28 | 110 |
| % del Total | 0,18% | 1,09% | 3,72% | 1,27% | 4,99% |

As we discussed, there are no hotels with the lowest score (1).
There are only 4 hotels in our sample with the minimum overall score (1.5), representing less than 1% of all hotels

There are only 28 hotels in our sample with an overall score between 1.5 and 2 which represents only 1.27% of all hotels

There are 110 hotels in our sample with the lowest overall score between 1.5 and 2.5 representing 4.99% of all hotels

Table 19: Low scores

| Nº habitaciones (Clases) | Puntuaciones Bajas (1,5 - 2 - 2,5) |
|---|---|
| 1-30 | 1,83% |
| 31-48 | 4,39% |
| 49-65 | 3,81% |
| 66-88 | 2,65% |
| 89-111 | 10,59% |
| 112-144 | 6,25% |
| 145-185 | 4,59% |
| 186-233 | 3,58% |
| 234-312 | 4,55% |
| 313-1468 | 7,76% |
| Total | 110 |
| Total | 4,99% |

Figure 10: Number of rooms and Low Ratings (1,5 - 2 - 2,5)
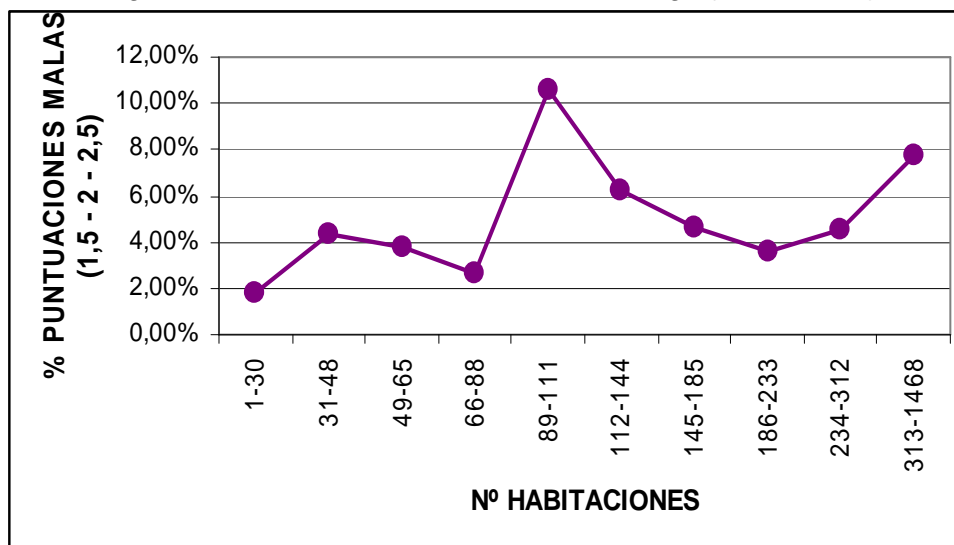


While only 1.83% of the smaller hotels (1 to 30 rooms) have bad scores (1,5 - 2 - 2,5) in the larger hotels (from 313 to 1468 rooms) this figure rises to 7.76%

We highlight the hotel segment between 89 and 111 rooms with a level of hotels with worse scores than the rest (10.59%)

Therefore a relationship is shown between the size of the establishment (measured by number of rooms) and the percentage of hotels with bad scores.

As the size of the hotel increases, there is a tendency that the proportion of hotels with the lowest scores increases.

<u>"Normal" Ratings (3 -3.5)</u>

The percentage of "Normal" ratings (3 - 3,5) also seems to have a certain degree of relation to the size of the establishments.

Table 20: Number of rooms and "Normal" Ratings (3 - 3,5)

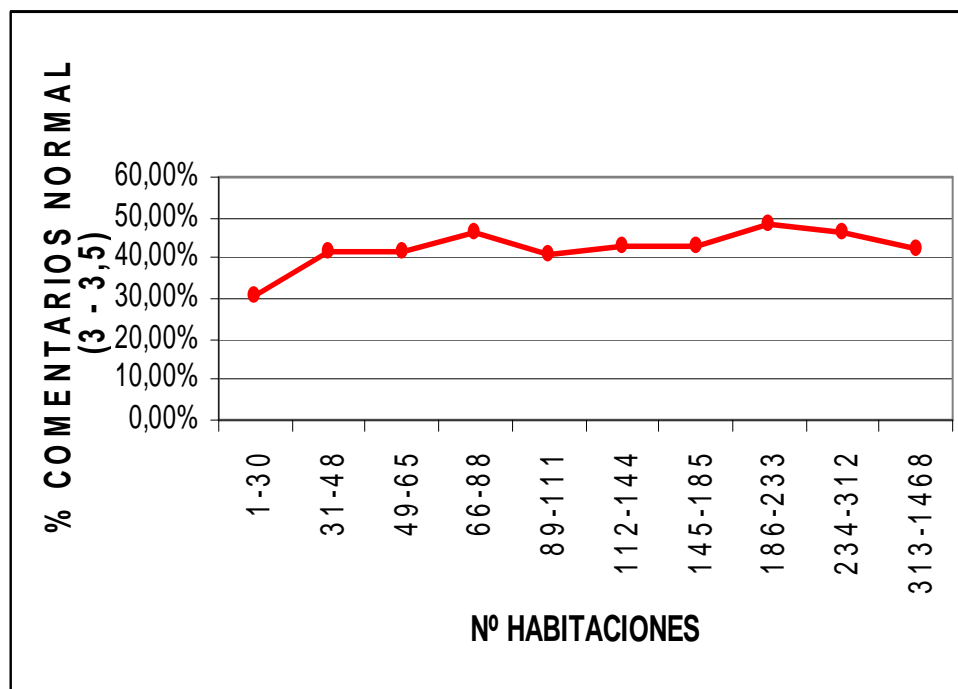| Rooms | Normal Ratings (3 – 3,5) Horizontal % | | |
|---|---|---|---|
| | 3 | 3,5 | 3 – 3,5 |
| 1-30 | 9,13% | 21,46% | 30,59% |
| 31-48 | 10,96% | 30,70% | 41,66% |
| 49-65 | 15,71% | 25,71% | 41,42% |
| 66-88 | 15,04% | 31,42% | 46,46% |
| 89-111 | 11,52% | 29,49% | 41,01% |
| 112-144 | 14,73% | 28,13% | 42,86% |
| 145-185 | 13,30% | 29,36% | 42,66% |
| 186-233 | 17,86% | 30,80% | 48,66% |
| 234-312 | 12,27% | 34,09% | 46,36% |
| 313-1468 | 17,81% | 24,66% | 42,47% |



Figure 11: Number of rooms and "Normal" Ratings (3 - 3,5)

As the size of the hotel increases, there is an upward trend in the proportion of hotels rated as "normal"

<u>Differences in rating</u>

Although having a high or low percentage of positive or negative comments is an interesting fact,

this does not provide all the information needed to draw conclusions.

If a hotel has a total of 50% of good reviews, this is not an indicative figure until it is compared with the percentage of bad reviews it has.
One hotel may have 60% of good comments and the rest average comments, while another hotel with the same percentage of good comments may have a remainder of bad comments.

To answer these questions we created three variables that measure the difference between the percentages of good and bad comments.

As we can see in the table below there are some negative values, indicating in these cases a higher percentage of high scores compared with the percentage of low scores. Although not usual, it is given twice in cases of extreme difference (5) - (1.5) for hotels from 89 to 111 rooms and hotels from 186 to 233 rooms

Also in these cases the differences decrease as the size of the hotels increases.

Table 21: Differences between High and Low Ratings

| Rooms. | Differences between High and Low Ratings (Horizontal %) | | |
|---|---|---|---|
| | *(5) - (1,5)* | *(5 - 4,5) - (1,5 - 2)* | *(5 - 4,5 - 4) - (1,5 - 2 - 2,5)* |
| 1-30 | 6,39% | 33,33% | 65,75% |
| 31-48 | 0,44% | 16,23% | 49,56% |
| 49-65 | 0,48% | 13,82% | 50,96% |
| 66-88 | 0,44% | 16,81% | 48,23% |
| 89-111 | -0,46% | 9,68% | 37,79% |
| 112-144 | 0,00% | 11,16% | 44,65% |
| 145-185 | 0,46% | 14,22% | 48,16% |
| 186-233 | -0,45% | 11,15% | 44,18% |
| 234-312 | 0,00% | 10,91% | 44,54% |
| 313-1468 | 0,00% | 6,85% | 42,01% |

The following graphs illustrate the behavior of these variables
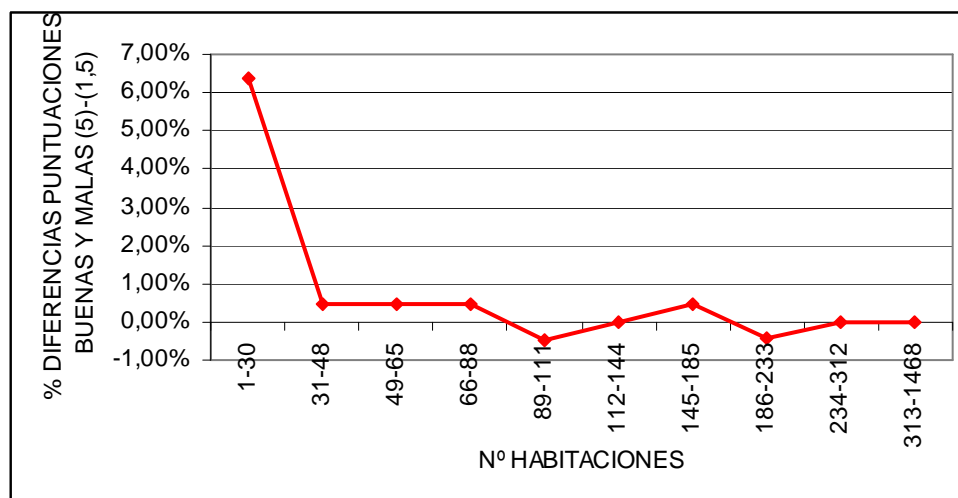


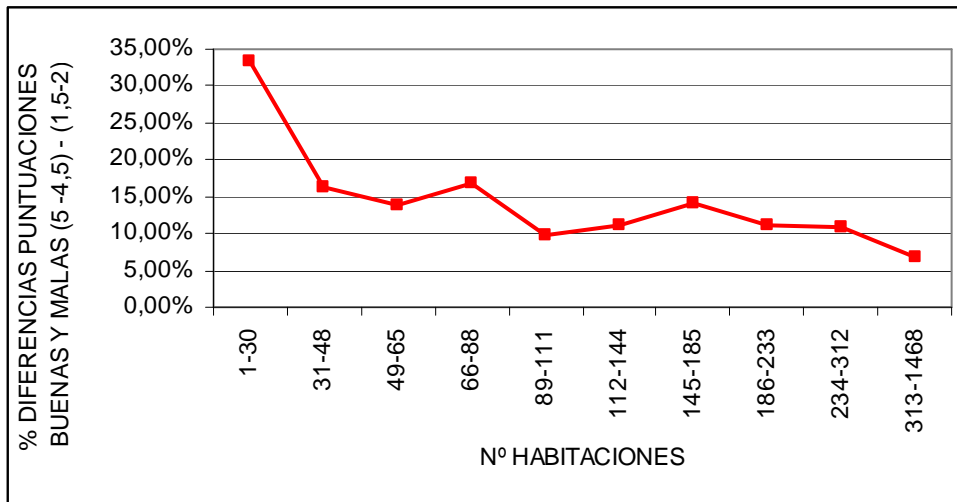Figure 12: Differences between High and Low Ratings (5) - (1.5)

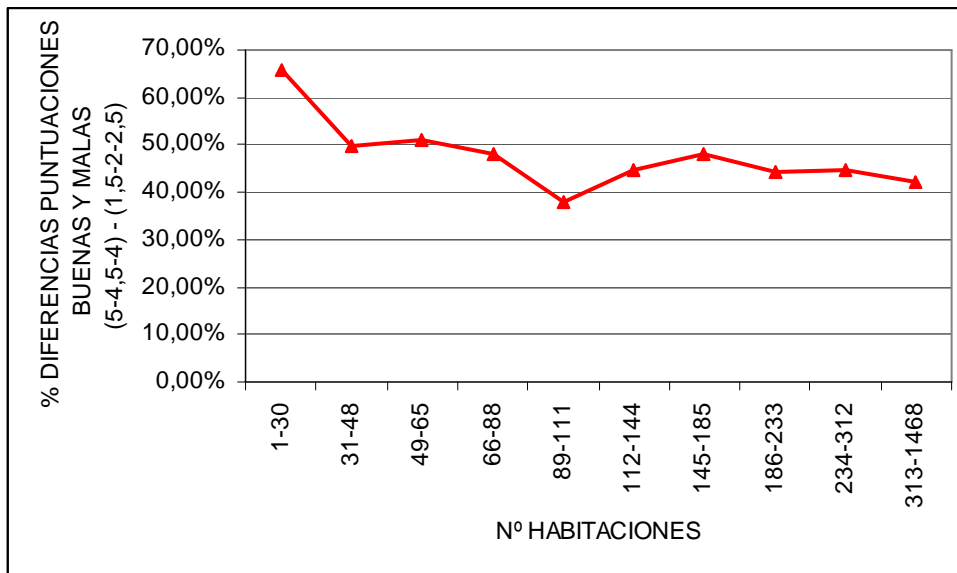Figure 13: Differences between High and Low Ratings (5 - 4.5) - (1.5 - 2)



Figure 14: Differences between High and Low Ratings (5 - 4.5 - 4) - (1,5 - 2 - 2.5)

As you increase the size of hotels, the difference between the percentage of hotels with the highest scores and the percentage of hotels with the lowest also reduces.

It is appreciated that there is a group of average-sized hotels (from 89 to 111 rooms) with minor differences in their scores (and in some cases to be negative)

## 6.1.3 The NCNH ratio and Rating

We believe that an important thing in the analysis of the comments about hotels is precisely the total number of these comments a hotel receives. TripAdvisor is very relevant to this figure because they understand that when there are a large number of comments, fraudulent reviews are less able to influence the average overall rating of the hotel. It is also a fact that affects the decision making process of the customer, and finally influences the position of the hotel in the

ranking.

We have analysed this variable but we think it is most relevant to analyse the ratio of number of comments and number of rooms, and we call this the NCNH Ratio (No. of Comments / Nª room) because we understand that you should not assess 300 reviews in a 300-room hotel in one review in a 60-room hotel.  In the first case we would have an NCNH 1 while the second there could be an NCNH 5, the ratio indicates that in the second hotel customers are 5 times more active than in the first. With the same number of comments, the first hotel room has a comment while the second has 5, thus the influence of size on the number of comments is eliminated.

The main measures of this ratio are:

Table 22: Ratio NCNH (Statistics)

| Variable | Number of comments / Number of Rooms |
|---|---|
| N | 2205 |
| Media | 2,06 |
| Desv Std | 3,15 |
| Variance | 9,95 |
| Mínimum | 0,07 |
| Máximum | 64,1 |
| Median | 1,233 |

Therefore we can see hotels whose relationship is almost zero while there are hotels where for each room there are over 60 comments.

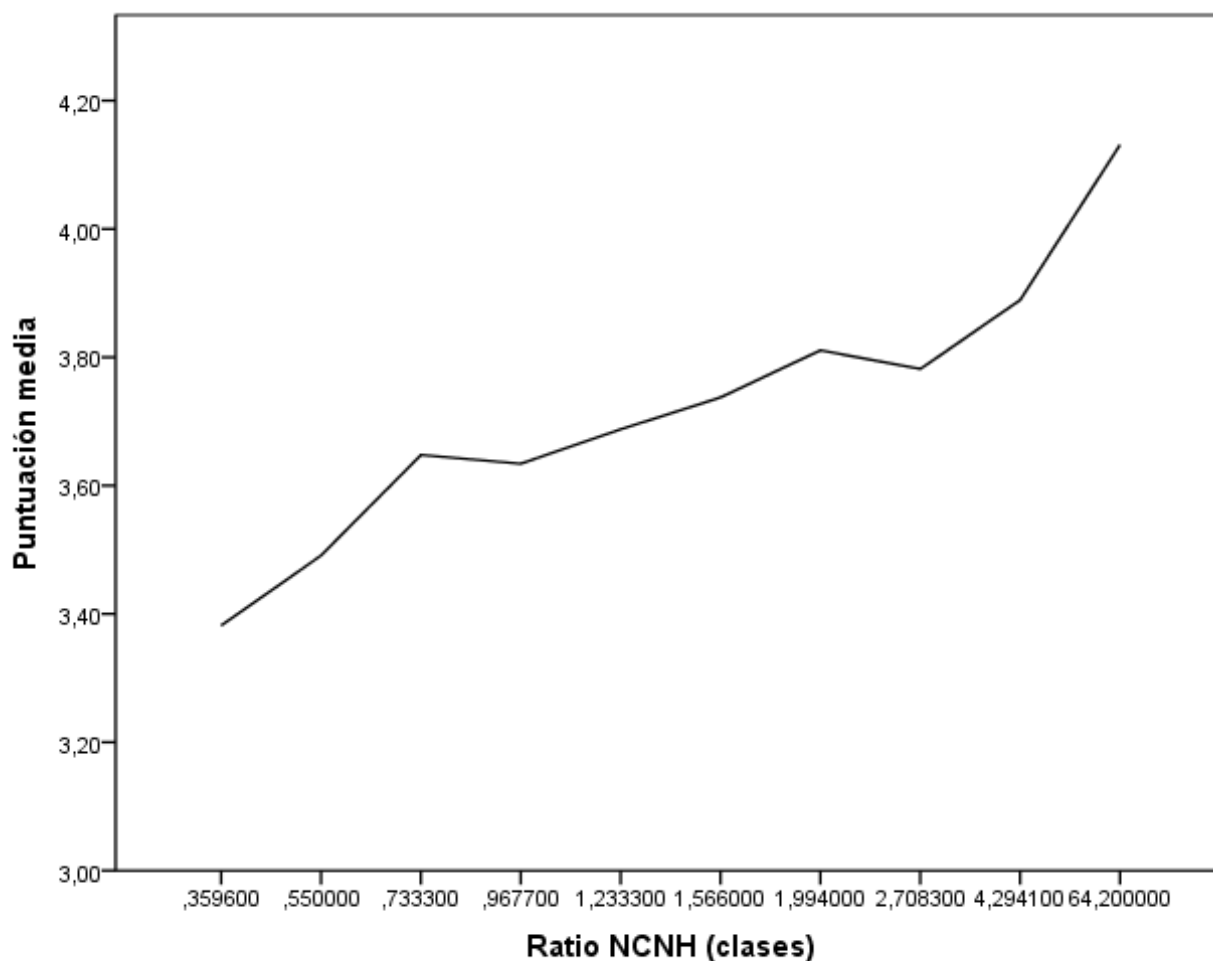The following graph illustrates this relationship.

Figure 15: NCNH and Rating

There is a significant relationship between these two variables so that as the value of the NCNH ratio increases, so does the average score of the hotel.

We will proceed in the same direction as the number of rooms and will conduct an analysis of variance:

First we split the observations into 10 homogeneous intervals.

Table 23: NCNH  Ratio (classes) and average score

| Descriptors | | | | |
|---|---|---|---|---|
| Rating | | | | |
| | N | Average | Typical Deviation | Typical error |
| ,359600 | 220 | 3,3818 | ,64965 | ,04380 |
| ,550000 | 221 | 3,4910 | ,54139 | ,03642 |
| ,733300 | 220 | 3,6477 | ,51661 | ,03483 |
| ,967700 | 220 | 3,6341 | ,54812 | ,03695 |
| 1,233300 | 221 | 3,6878 | ,53282 | ,03584 |
| 1,566000 | 219 | 3,7374 | ,52310 | ,03535 |
| 1,994000 | 222 | 3,8108 | ,49456 | ,03319 |
| 2,708300 | 220 | 3,7818 | ,51029 | ,03440 |
| 4,294100 | 221 | 3,8891 | ,47927 | ,03224 |

| | | | | | |
|---|---|---|---|---|---|
| 64,200000 | 221 | 4,1312 | ,45324 | ,03049 |
| Total | 2205 | 3,7195 | ,56206 | ,01197 |
| Model Fixed effects | | | ,52720 | ,01123 |
| Variable Effects | | | | ,06590 |

As we can see, the average score for each interval grows with increasing NCNH ratio value.

There is a tendency for the NCNH ratio to decrease as the size of the hotels increases.

## 6.2 ANALYSIS OF COASTAL TOURISTIC AREAS OF SPAIN AND PORTUGAL

Then we made analyses of the hotels located in coastal destinations in Spain and southern Portugal from the information in our database.

In this case we grouped the hotels by Tourism Zones, comparing key indicators of these 14 destinations.

### 6.2.1 According to the mean score of the hotels.

The following table shows the average values of the average overall rating for each destination of hotels shown.

Table 31: Rating by touristic áreas

| | | Rating | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recount | Average | Median | Mode | Máximum | Mínimum |
| destino1 | Algarve | 147 | 3,76 | 3,50 | 3,50 | 4,50 | 2,50 |
| | Costa de Almería | 22 | 3,18 | 3,25 | 3,50 | 4,00 | 2,00 |
| | Costa Blanca | 187 | 3,69 | 3,50 | 4,00 | 5,00 | 2,00 |
| | Costa Brava | 138 | 3,50 | 3,50 | 4,00 | 4,50 | 2,00 |
| | Costa de Barcelona | 413 | 3,82 | 4,00 | 4,00 | 5,00 | 2,00 |
| | Costa de Castellón | 31 | 3,60 | 3,50 | 3,00 | 4,50 | 3,00 |
| | Costa de la Luz | 80 | 3,78 | 4,00 | 4,00 | 5,00 | 2,00 |
| | Costa de Murcia | 61 | 3,80 | 4,00 | 4,00 | 5,00 | 2,50 |
| | Costa de Valencia | 36 | 3,40 | 3,50 | 3,50 | 4,00 | 3,00 |
| | Costa del Sol | 233 | 3,74 | 4,00 | 4,00 | 5,00 | 1,50 |
| | Costa Dorada | 123 | 3,52 | 3,50 | 4,00 | 5,00 | 1,50 |
| | Costa Tropical | 18 | 3,72 | 3,75 | 3,50 | 4,50 | 2,50 |
| | Islas Baleares | 385 | 3,74 | 4,00 | 4,00 | 5,00 | 2,00 |
| | Islas Canarias | 337 | 3,78 | 4,00 | 4,00 | 5,00 | 2,00 |

We note that the hotels on the Costa de Barcelona have the highest mean score (PGM) (3.82), while the lowest score lies with the hotels on the coast of Almeria (3.18).

The following chart provides the most information about the data:
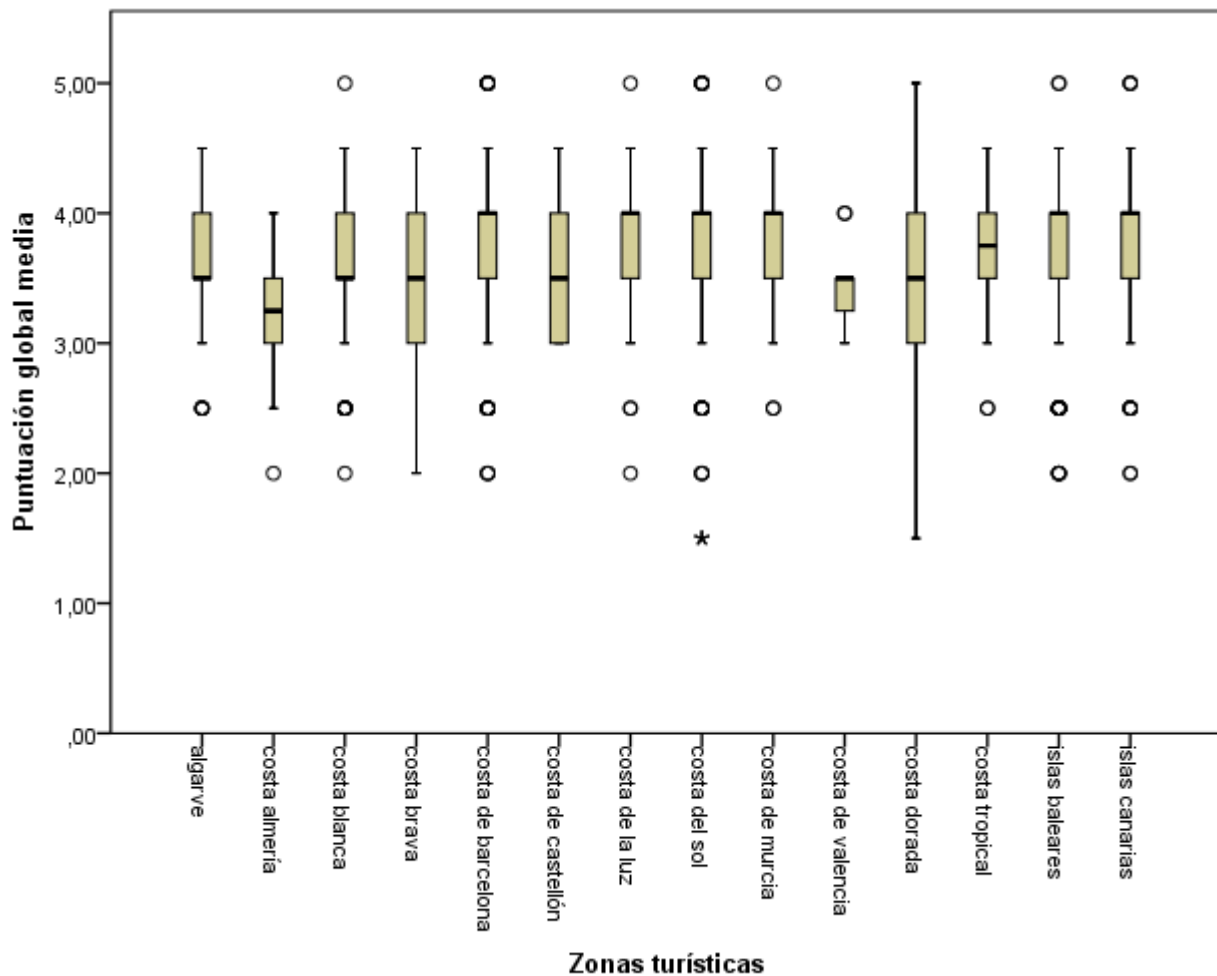


Graph 17: Average Score for touristic areas

The Gold Coast is the one with more extreme values. The two locations with lower minimum values are the Costa del Sol and Costa Dorada, both have hotels with overall average scores of 1.5. The Coast of Castellón and Valencia have the highest minimum for any hotel in these destinations. However the Costa de Valencia which together with the Costa de Almería have lower highs than any hotel has on these shores.

## 6.2.2 According to the average size of hotels (measured by number of rooms)

The following table shows the average values ??of the size of the hotels (measured by number of rooms) for each destination.

Table 32: Number of rooms for touristic areas

|  |  | Number of rooms | | | | | |
|  |  | Recount | Average | Median | Mode | Máximum | Mínimum |
|---|---|---|---|---|---|---|---|
| destino1 | Algarve | 147 | 130,89 | 104,00 | 60,00 | 627,00 | 10,00 |
|  | Costa de Almería | 22 | 159,73 | 119,00 | 15,00 | 549,00 | 15,00 |
|  | Costa Blanca | 187 | 155,17 | 123,00 | 52,00 | 794,00 | 7,00 |
|  | Costa Brava | 138 | 155,76 | 119,50 | 208,00 | 870,00 | 11,00 |
|  | Costa de Barcelona | 413 | 107,34 | 78,00 | 40,00 | 601,00 | 9,00 |
|  | Costa de Castellón | 31 | 169,77 | 154,00 | 10,00 | 564,00 | 10,00 |
|  | Costa de la Luz | 80 | 114,60 | 53,50 | 15,00 | 1200,00 | 5,00 |
|  | Costa de Murcia | 61 | 113,30 | 100,00 | 100,00 | 449,00 | 2,00 |
|  | Costa de Valencia | 36 | 101,56 | 87,00 | 51,00 | 319,00 | 10,00 |
|  | Costa del Sol | 233 | 149,35 | 126,00 | 9,00 | 650,00 | 5,00 |
|  | Costa Dorada | 123 | 203,38 | 163,00 | 30,00 | 968,00 | 7,00 |
|  | Costa Tropical | 18 | 118,39 | 97,50 | 35,00 | 325,00 | 6,00 |
|  | Islas Baleares | 385 | 155,92 | 132,00 | 36,00 | 1468,00 | 5,00 |
|  | Islas Canarias | 337 | 198,99 | 167,50 | 100,00 | 1136,00 | 9,00 |

The values in the table indicate that the hotels on the Costa Dorada are on average the largest by number of rooms (203.38 rooms), while the smallest are the Costa de Valencia (101 rooms)
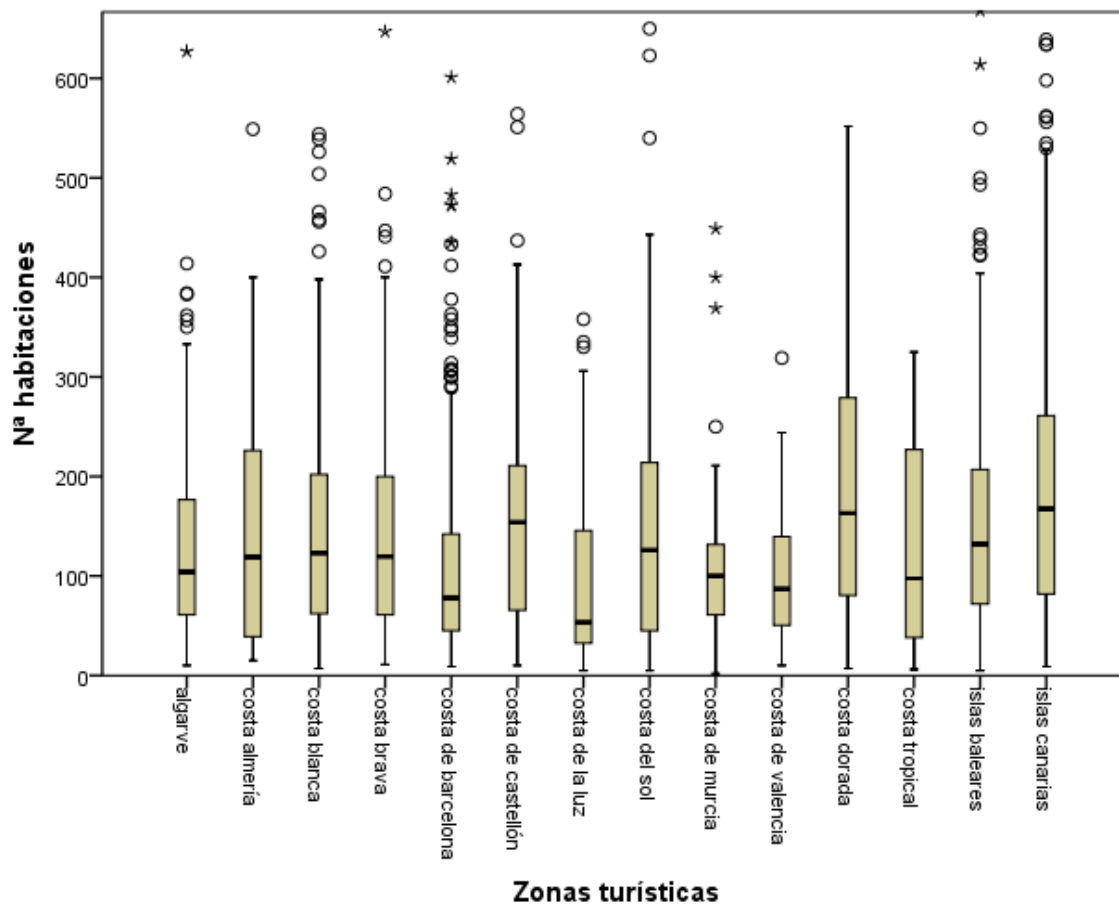
Figure 18: Average Number of rooms for touristic areas

In all locations there are very small hotels (includso 2 rooms). Surprisingly, the Costa de Valencia has hotels with over 319 rooms.

## 6.2.3 According to the average number of comments about hotels.

The following table shows the average values ?of the number of reviews about hotels for each destination.

Table 33: Number of comments in touristic areas

| | | Nº Commentaries | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recount | Mean | Median | Mode | Máximum | Mínimum |
| destino1 | Algarve | 147 | 192,24 | 119,00 | 46,00 | 1135,00 | 20,00 |
| | Costa de Almería | 22 | 80,32 | 62,50 | 24,00 | 253,00 | 20,00 |
| | Costa Blanca | 187 | 230,67 | 119,00 | 24,00 | 1755,00 | 20,00 |
| | Costa Brava | 138 | 123,94 | 75,50 | 88,00 | 1166,00 | 21,00 |
| | Costa de Barcelona | 413 | 304,45 | 207,00 | 95,00 | 2188,00 | 20,00 |
| | Costa de Castellón | 31 | 74,00 | 44,00 | 35,00 | 255,00 | 20,00 |
| | Costa de la Luz | 80 | 112,18 | 52,00 | 42,00 | 840,00 | 20,00 |
| | Costa de Murcia | 61 | 71,95 | 54,00 | 23,00 | 339,00 | 20,00 |
| | Costa de Valencia | 36 | 71,17 | 66,50 | 31,00 | 153,00 | 20,00 |
| | Costa del Sol | 233 | 227,80 | 176,00 | 39,00 | 987,00 | 20,00 |
| | Costa Dorada | 123 | 185,41 | 104,00 | 37,00 | 917,00 | 20,00 |
| | Costa Tropical | 18 | 125,72 | 67,00 | 54,00 | 530,00 | 20,00 |
| | Islas Baleares | 385 | 187,16 | 113,00 | 20,00 | 2994,00 | 20,00 |
| | Islas Canarias | 337 | 303,25 | 186,00 | 22,00 | 1779,00 | 20,00 |

In this case the hotels on the Costa de Barcelona and in the Canary Islands are on average receiving more comments (304 and 303 respectively), while the Costa de Valencia and Murcia Coast have had the least number of comments received (71, 17 and 71.95)

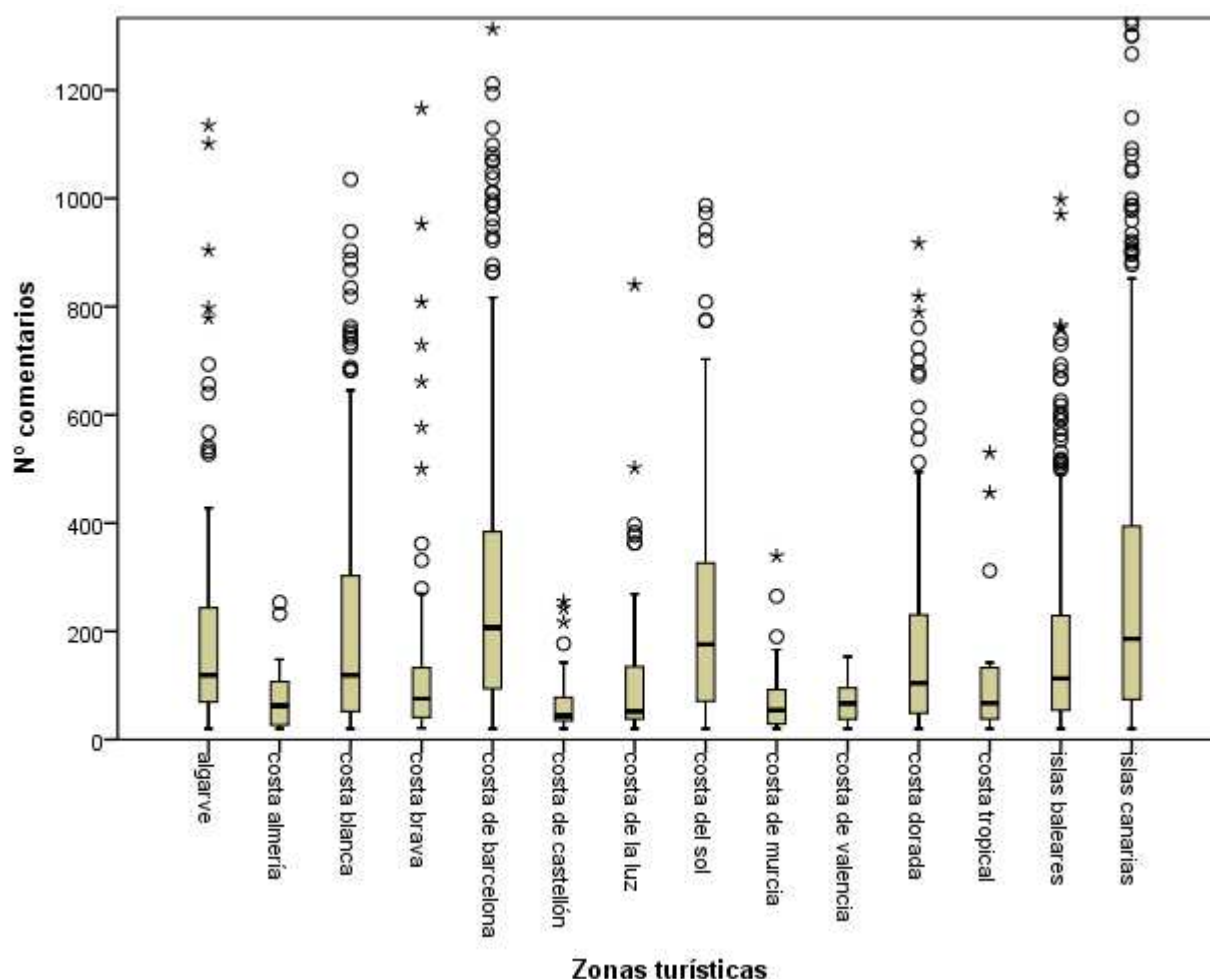The following chart provides the most information about the data:

Figure 19: Number of reviews for touristic areas

All the destinations have hotels with the minimum value of comments (20). And the Balearic Islands and Costa Barcelona hotels have over 2,000 comments. Surprisingly, no hotels have more than 153 comments on the coast of Valencia.

## 6.2.4 According to the average value of the NCNH ratio (No. Comments / Bedroom) on its hotels.

The following table shows the average values ??of the NCNH ratio (No. comments / Number of rooms) hotels for each destination

Table 34: Ratio for touristic areas NCNH Ratio (Nº of Comentaries / Nº of Rooms)

|  |  | NCNH Ratio (Nº of Comentaries / Nº of Rooms) | | | | | |
|---|---|---|---|---|---|---|---|
|  |  | Recount | Average | Median | Mode | Máximum | Mínimum |
| destino1 | Algarve | 147 | 1,78 | 1,38 | 1,00 | 13,08 | ,07 |
|  | Costa de Almería | 22 | ,86 | ,55 | ,19 | 3,26 | ,19 |
|  | Costa Blanca | 187 | 1,74 | 1,15 | ,60 | 24,35 | ,09 |
|  | Costa Brava | 138 | ,97 | ,62 | ,56 | 8,24 | ,11 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Costa de Barcelona | 413 | 4,04 | 2,83 | 2,11 | 64,10 | ,13 |
| Costa de Castellón | 31 | ,60 | ,43 | ,13 | 2,50 | ,13 |
| Costa de la Luz | 80 | 1,76 | 1,13 | ,15 | 24,80 | ,15 |
| Costa de Murcia | 61 | ,95 | ,63 | ,21 | 10,00 | ,15 |
| Costa de Valencia | 36 | 1,14 | ,70 | 1,00 | 4,70 | ,15 |
| Costa del Sol | 233 | 2,34 | 1,50 | ,70 | 33,60 | ,09 |
| Costa Dorada | 123 | 1,10 | ,75 | ,50 | 10,43 | ,07 |
| Costa Tropical | 18 | 2,00 | ,87 | ,13 | 13,00 | ,13 |
| Islas Baleares | 385 | 1,48 | 1,02 | 1,00 | 11,83 | ,14 |
| Islas Canarias | 337 | 1,75 | 1,30 | 1,00 | 45,70 | ,10 |

The hotels on the Costa de Barcelona have received on average more comments in proportion to their size. The hotels on the Costa del Sol are in second place.

The following chart provides the most information about the data:
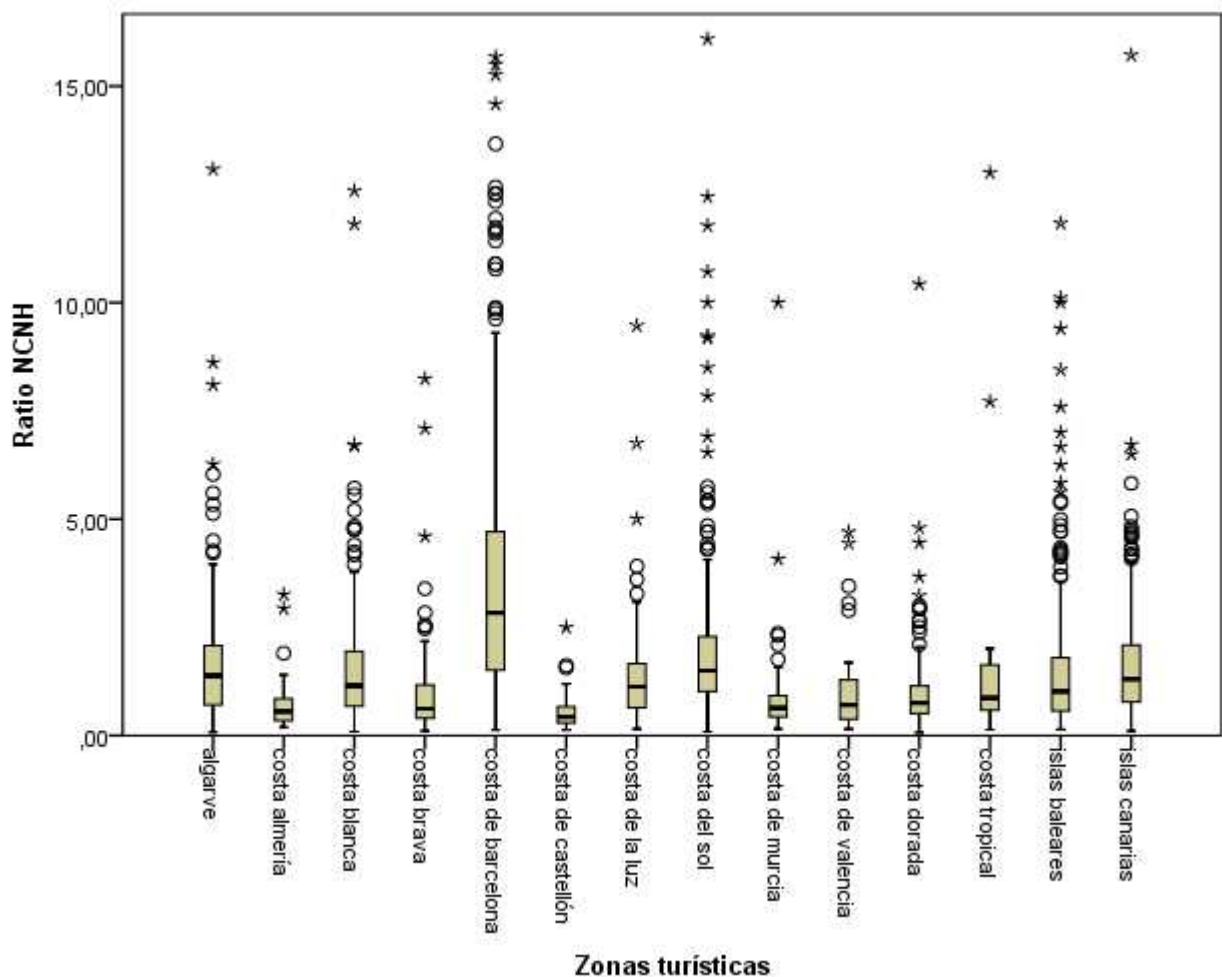


Figure 20: NCNH Ratio for tourist areas

The Costa de Barcelona is especially outstanding in this section by far over other destinations, finding hotels with a ratio of more than 64 comments per room. The hotels on the Costa Almeria, Costa de Valencia, Costa Castellón are characterized by less active clients when posting comments on TripAdvisor.

Comparing this figure with the number of comments, we recognize the utility of the NCNH ratio. Concerning the number of views, the Canary Islands is on a similar level to Barcelona, however, using the NCNH Ratio, we can observe a difference between destinations, which can be interpreted as that on average most comments are written on the hotels Barcelona (proportionally of the same size). A similar analysis was made comparing the Costa del Sol, Costa Dorada and Balearic Islands.

## 6.2.5 According to the mean percentage of "Excellent" and "Very Good" comments

The following table shows the average values of the percentage of "Excellent" and "Very good" comments about hotels for each destination.

Table 35: Percentage of "Excellent" and "Very Good" comments for tourist areas.

| | | % Excellent + Very Good comments | | | | |
|---|---|---|---|---|---|---|
| | | Recount | Average | Median | Mode | Máximum | Mínimum |
| destino1 | Algarve | 147 | ,67 | ,68 | ,43 | ,96 | ,22 |
| | Costa de Almería | 22 | ,47 | ,49 | ,13 | ,80 | ,13 |
| | Costa Blanca | 187 | ,64 | ,66 | ,72 | ,98 | ,10 |
| | Costa Brava | 138 | ,59 | ,63 | ,67 | 1,00 | ,13 |
| | Costa de Barcelona | 413 | ,70 | ,74 | ,50 | ,99 | ,11 |
| | Costa de Castellón | 31 | ,63 | ,62 | ,34 | ,96 | ,34 |
| | Costa de la Luz | 80 | ,68 | ,69 | ,43 | ,98 | ,20 |
| | Costa de Murcia | 61 | ,68 | ,71 | ,21 | ,98 | ,21 |
| | Costa de Valencia | 36 | ,53 | ,54 | ,53 | ,80 | ,21 |
| | Costa del Sol | 233 | ,67 | ,71 | ,83 | ,99 | ,00 |
| | Costa Dorada | 123 | ,60 | ,62 | ,67 | ,99 | ,07 |
| | Costa Tropical | 18 | ,65 | ,65 | ,19 | ,94 | ,19 |
| | Islas Baleares | 385 | ,67 | ,71 | ,83 | ,98 | ,12 |
| | Islas Canarias | 337 | ,68 | ,71 | ,45 | ,98 | ,11 |

It is observed that the hotels on the Costa de Barcelona are receiving on average more "Excellent"

and "Very Good" comments. 70% of the comments received by half these hotels are in the "Excellent" or "Very Good" category. The hotels on the Costa Almeria have received the lowest number of such comments (47%)

The following chart provides the most information about the data:
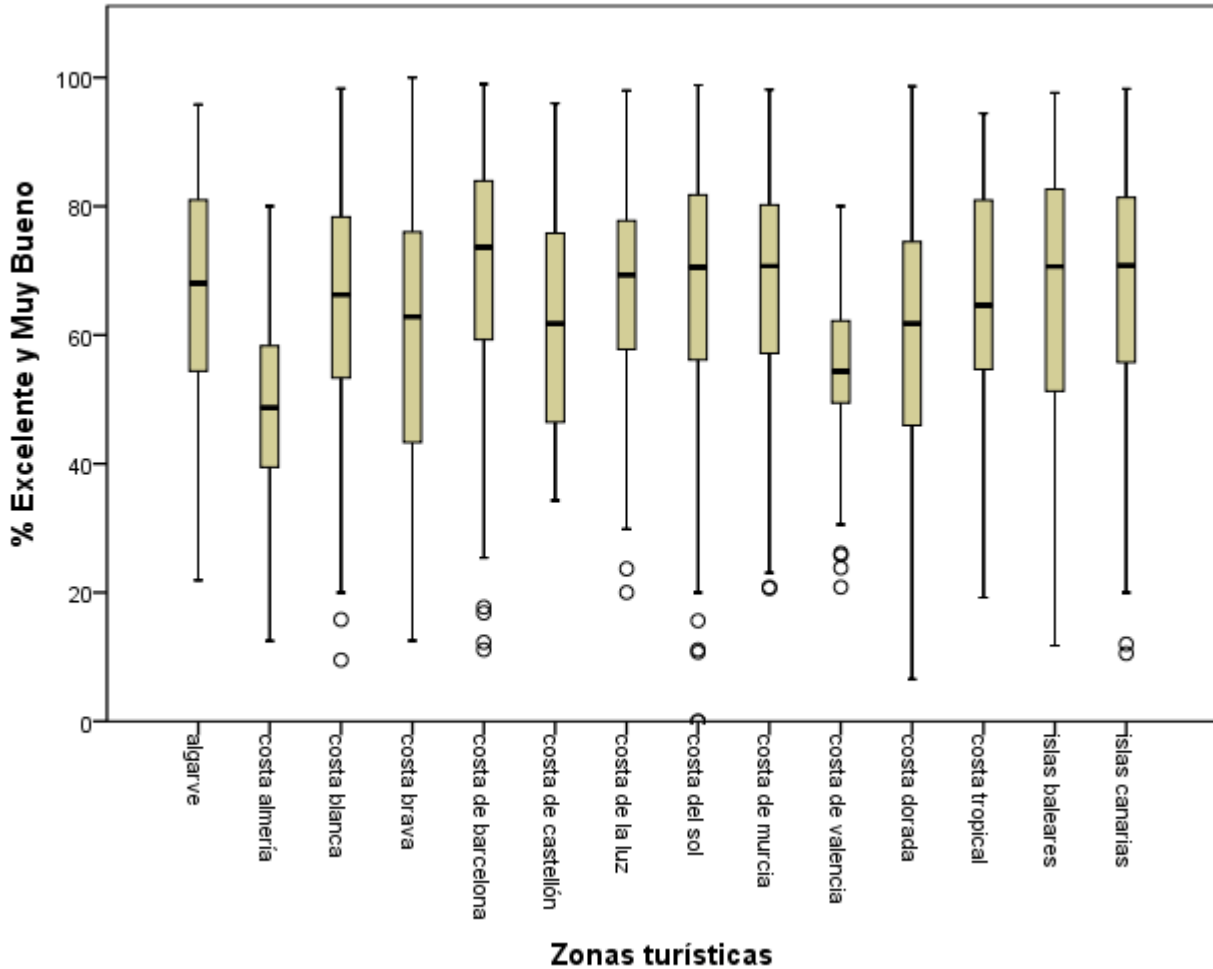


Figure 21: Percentage of "Excellent" and "Very Good" comments for tourist areas

We find some destinations that reach all possible outcomes. So the Costa del Sol has both "near perfect" hotels where all comments qualify them as excellent or very good, but also the hotels where no customer has described them this way.

## 6.2.6 According to the mean percentage of "Bad" and "Terrible" comments

The following table shows the average values ??of the percentage of "Bad" and "Terrible" comments about hotels for each destination

Table 36: Percentage of "Bad" and "Terrible" comments for tourist areas

| | | % Bad + Very Bad comments | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recount | Average | Median | Mode | Máximum | Mínimum |
| destino1 | Algarve | 147 | ,14 | ,10 | ,00 | ,58 | ,00 |
| | Costa de Almería | 22 | ,29 | ,24 | ,00 | ,75 | ,00 |
| | Costa Blanca | 187 | ,16 | ,12 | ,00 | ,67 | ,00 |
| | Costa Brava | 138 | ,21 | ,17 | ,00 | ,68 | ,00 |
| | Costa de Barcelona | 413 | ,12 | ,09 | ,00 | ,65 | ,00 |
| | Costa de Castellón | 31 | ,16 | ,12 | ,00 | ,41 | ,00 |
| | Costa de la Luz | 80 | ,14 | ,12 | ,00 | ,72 | ,00 |
| | Costa de Murcia | 61 | ,13 | ,11 | ,00 | ,60 | ,00 |
| | Costa de Valencia | 36 | ,20 | ,18 | ,33 | ,51 | ,00 |
| | Costa del Sol | 233 | ,15 | ,11 | ,00 | ,91 | ,00 |
| | Costa Dorada | 123 | ,20 | ,16 | ,00 | ,83 | ,00 |
| | Costa Tropical | 18 | ,15 | ,11 | ,00 | ,54 | ,00 |
| | Islas Baleares | 385 | ,16 | ,13 | ,00 | ,70 | ,00 |
| | Islas Canarias | 337 | ,14 | ,12 | ,00 | ,72 | ,00 |

The hotels on the Costa Almeria are receiving on average more "Terrible" and "Bad" comments. 29% of the comments received about half these hotels are in the "Terrible" or "Bad" category, while the Costa de Barcelona received the least number of such comments (12%)

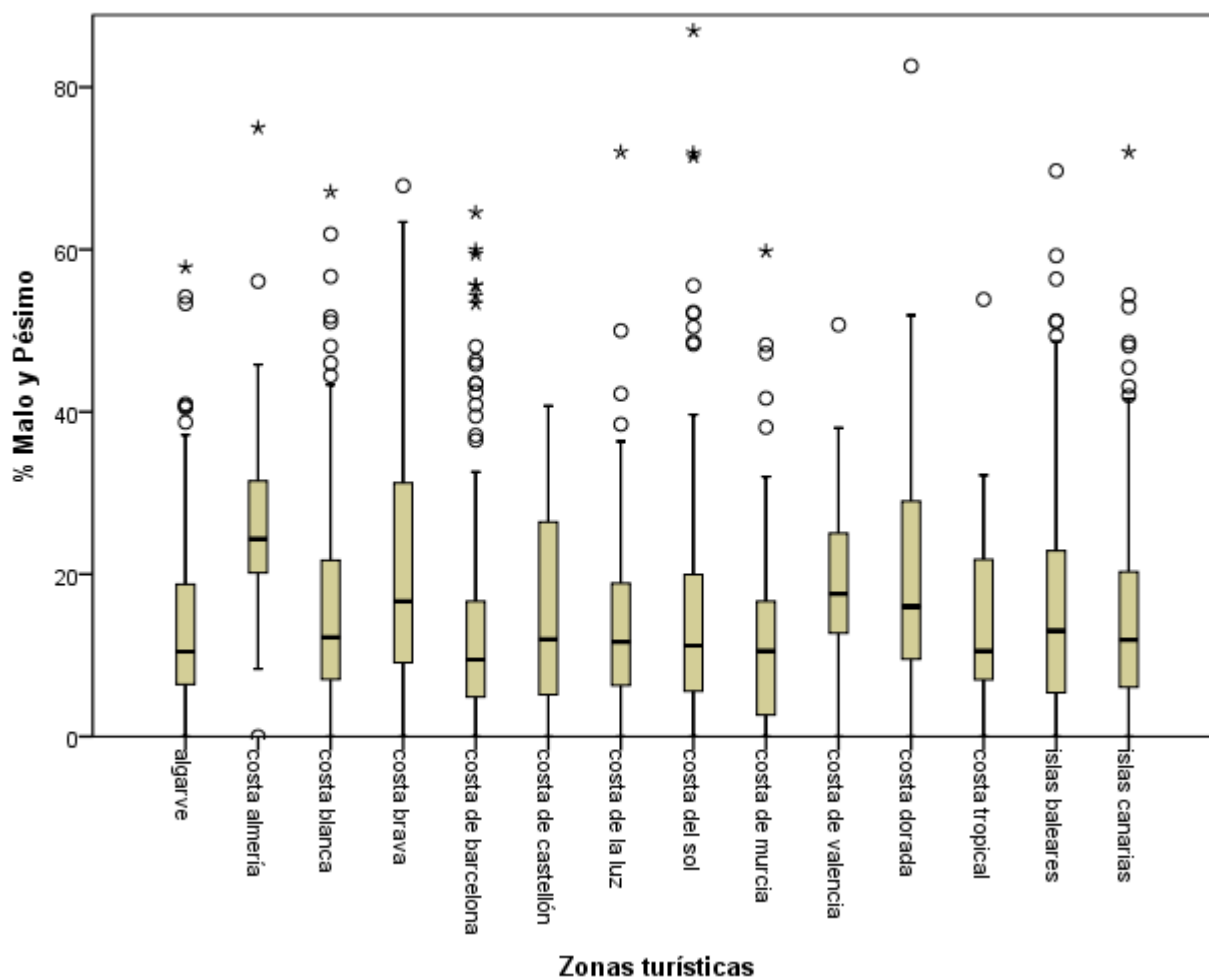The following chart provides the most information about the data:

Figure 22: Percentage of "Bad" and "Terrible" comments for touristic areas

The Costa del Sol returns to cover all possible valuations. In line with the previous section there are hotels without any negative comments and hotels where over 90% of their customers value them as bad or very bad.

## 6.2.7 According to the mean percentage of "Family" customers

The following table shows the average values ?of the percentage of comments made by customers traveling in "Families"

Table 37: Percentage of "Family" comments for touristic areas

|  |  | Family comments % |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| destino1 | Algarve | 147 | ,33 | ,31 | ,05 | ,81 | ,05 |
|  | Costa de Almería | 22 | ,42 | ,42 | ,25 | ,89 | ,00 |
|  | Costa Blanca | 187 | ,31 | ,30 | ,50 | 1,00 | ,05 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Costa Brava | 138 | ,34 | ,32 | ,17 | ,87 | ,04 |
| Costa de Barcelona | 413 | ,22 | ,18 | ,13 | ,76 | ,00 |
| Costa de Castellón | 31 | ,47 | ,52 | ,29 | ,84 | ,08 |
| Costa de la Luz | 80 | ,26 | ,19 | ,05 | ,84 | ,03 |
| Costa de Murcia | 61 | ,23 | ,19 | ,10 | ,95 | ,02 |
| Costa de Valencia | 36 | ,26 | ,20 | ,00 | ,78 | ,00 |
| Costa del Sol | 233 | ,25 | ,21 | ,11 | ,86 | ,00 |
| Costa Dorada | 123 | ,46 | ,46 | ,14 | ,93 | ,05 |
| Costa Tropical | 18 | ,37 | ,26 | ,63 | ,82 | ,09 |
| Islas Baleares | 385 | ,30 | ,23 | ,00 | ,94 | ,00 |
| Islas Canarias | 337 | ,30 | ,28 | ,25 | ,80 | ,00 |

The hotels in the Costa Castellón on average have a higher percentage of "Family" customers (47%), the Barcelona Coast has fewer comments of this type received (22%)

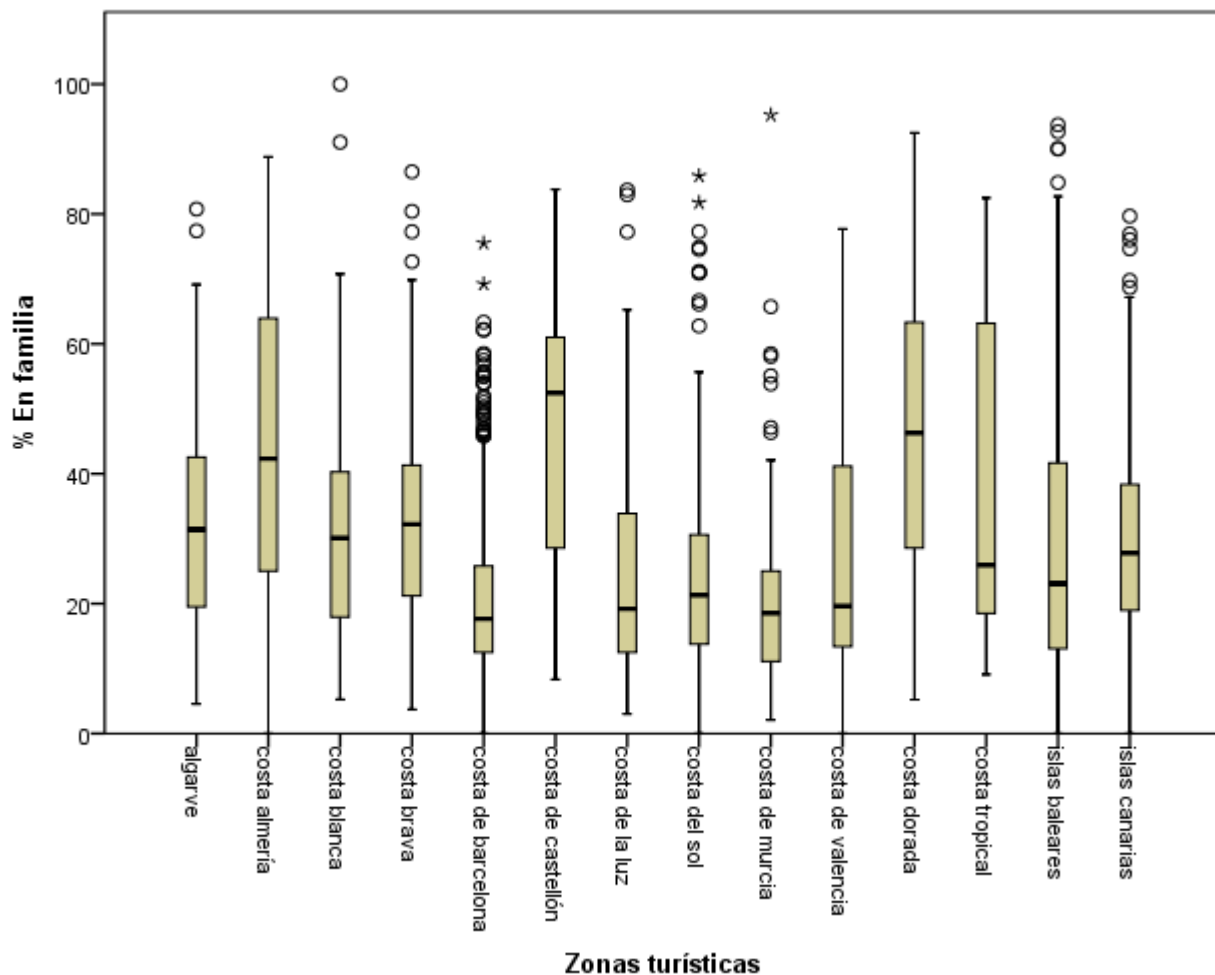The following chart provides the most information about the data:



Figure 23: Percentage of comments written by tourists traveling in "Family" for touristic areas

Found on the Costa Blanca hotels where all family customers (who have written a review on

TripAdvisor) have travelled. The Costa de Barcelona and Valencia are both less "family orientated" destinations on the basis of feedback from customers.

## 6.2.8 According to the mean percentage of "Couples" customers

The following table shows the average values ??of the percentage of comments made by customers travelling as "Couples"

Table 38: Percentage of "Couples" comments for touristic areas

| | | Couple comments % | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| destino1 | Algarve | 147 | ,53 | ,55 | ,33 | ,86 | ,15 |
| | Costa de Almería | 22 | ,40 | ,38 | ,09 | ,65 | ,09 |
| | Costa Blanca | 187 | ,49 | ,50 | ,50 | ,88 | ,00 |
| | Costa Brava | 138 | ,46 | ,45 | ,50 | ,82 | ,08 |
| | Costa de Barcelona | 413 | ,51 | ,51 | ,44 | ,85 | ,00 |
| | Costa de Castellón | 31 | ,41 | ,38 | ,09 | ,92 | ,09 |
| | Costa de la Luz | 80 | ,55 | ,57 | ,60 | ,88 | ,14 |
| | Costa de Murcia | 61 | ,51 | ,52 | ,51 | ,86 | ,00 |
| | Costa de Valencia | 36 | ,44 | ,44 | ,30 | ,77 | ,14 |
| | Costa del Sol | 233 | ,54 | ,55 | ,67 | ,92 | ,12 |
| | Costa Dorada | 123 | ,40 | ,39 | ,50 | ,79 | ,05 |
| | Costa Tropical | 18 | ,49 | ,55 | ,14 | ,74 | ,14 |
| | Islas Baleares | 385 | ,49 | ,49 | ,50 | 1,00 | ,00 |
| | Islas Canarias | 337 | ,56 | ,57 | ,50 | ,88 | ,18 |

We found that hotels in the Canary Islands which on average have a higher percentage of "Couples" customers (56%), the Costa Dorada and the Costa de Almería are those with the least (40%)

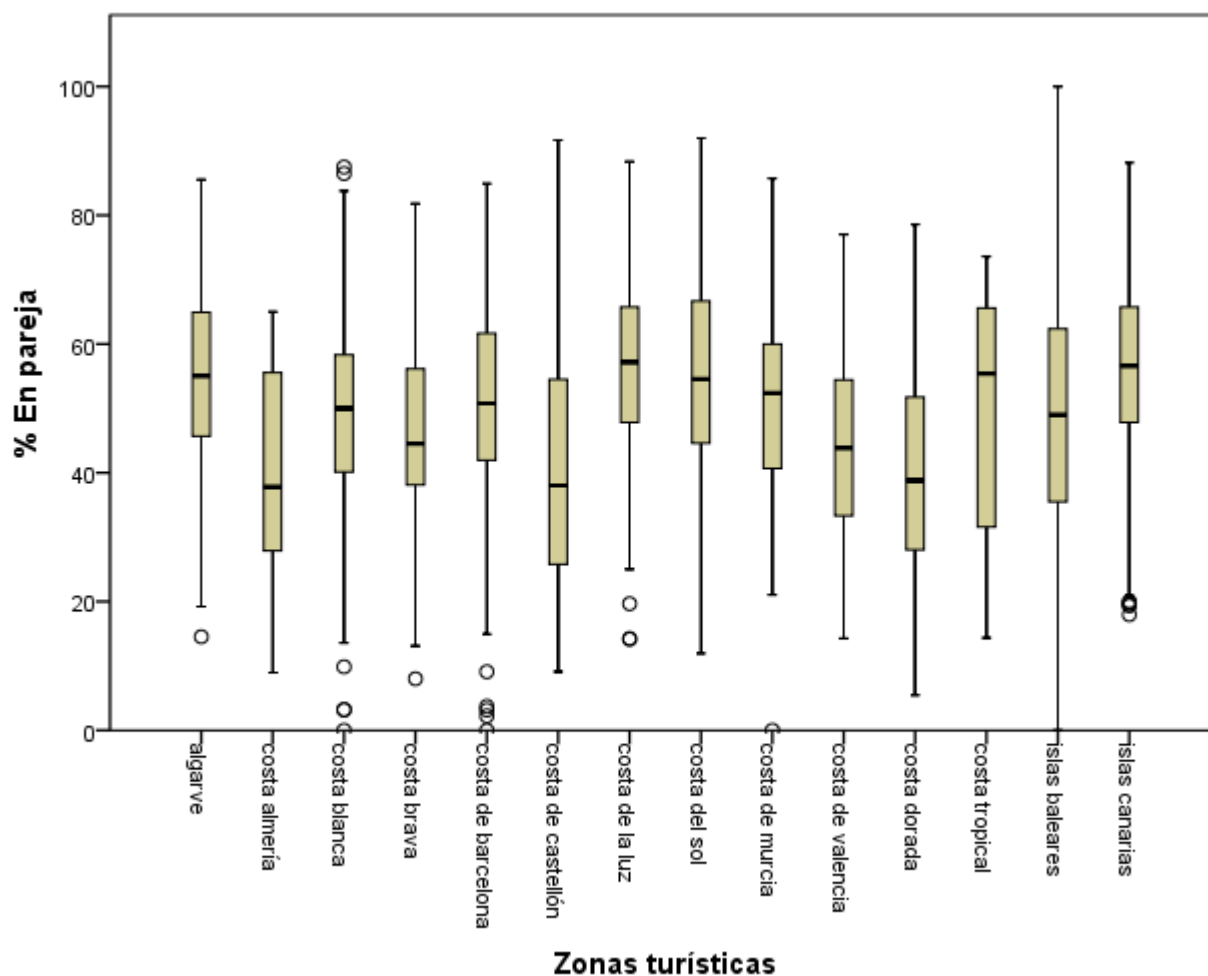The following chart provides the most information about the data:

Figure 24: Percentage of comments written by tourists traveling as "Couples" by touristic zones

In the Balearic Islands hotels where all customers (those who have written reviews on TripAdvisor) have been a couple but also hotels where nobody has travelled as a couple.

## 6.2.9 According to the mean percentage of "Single" customers

The following table shows the average values ?of the percentage of comments made by customers traveling alone

Table 39: Percentage of comments made by customers travelleing alone  for tourist areas

| | | "Alone" comments % | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| destino1 | Algarve | 147 | ,03 | ,02 | ,00 | ,23 | ,00 |
| | Costa de Almería | 22 | ,02 | ,01 | ,00 | ,17 | ,00 |
| | Costa Blanca | 187 | ,05 | ,03 | ,00 | ,38 | ,00 |
| | Costa Brava | 138 | ,03 | ,02 | ,00 | ,26 | ,00 |
| | Costa de Barcelona | 413 | ,07 | ,05 | ,00 | ,45 | ,00 |
| | Costa de Castellón | 31 | ,02 | ,01 | ,00 | ,13 | ,00 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Costa de la Luz | 80 | ,03 | ,03 | ,00 | ,22 | ,00 |
| Costa de Murcia | 61 | ,05 | ,04 | ,00 | ,20 | ,00 |
| Costa de Valencia | 36 | ,06 | ,04 | ,00 | ,29 | ,00 |
| Costa del Sol | 233 | ,05 | ,04 | ,00 | ,28 | ,00 |
| Costa Dorada | 123 | ,02 | ,01 | ,00 | ,14 | ,00 |
| Costa Tropical | 18 | ,02 | ,01 | ,00 | ,07 | ,00 |
| Islas Baleares | 385 | ,05 | ,03 | ,00 | ,39 | ,00 |
| Islas Canarias | 337 | ,05 | ,03 | ,00 | ,54 | ,00 |

In all destinations, the most repeated value, 0 in this section is generally made by a very low percentage of customers traveling "Solo" 7% being the maximum on the Barcelona Coast.

The following chart provides the most information about the data:
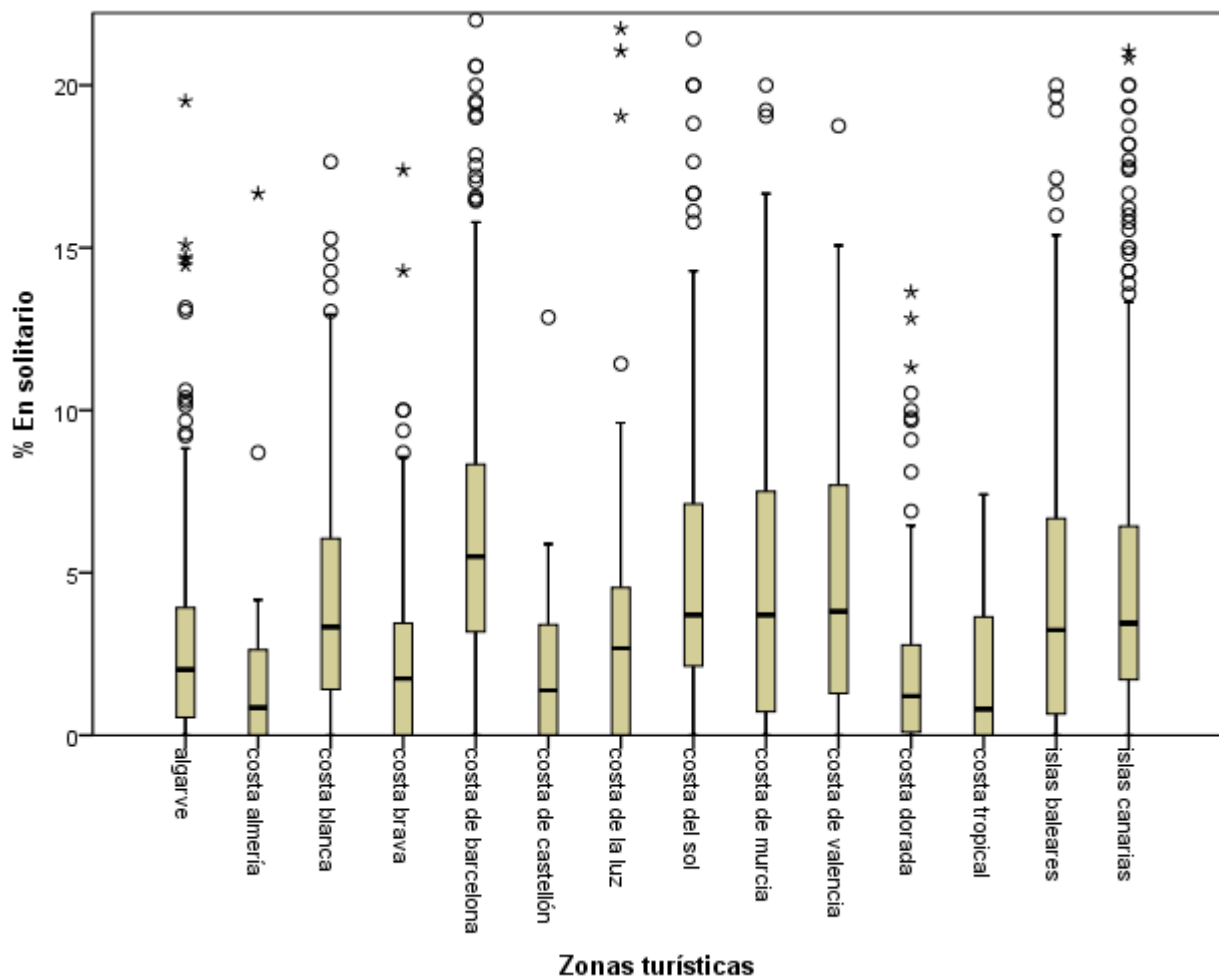


Figure 25: Percentage of comments written by tourists traveling "Solo" by touristic zones

In all destinations without exception there are hotels where no customer has gone solo.

## 6.2.10 According to the mean percentage of "Business" customers

The following table shows the average values ?of the percentage of comments made by customers traveling for "Business"

Table 40: Percentage of "Business" comments in tourist areas

| | | "On Business" comments % | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| destino1 | Algarve | 147 | ,02 | ,01 | ,00 | ,21 | ,00 |
| | Costa de Almería | 22 | ,03 | ,01 | ,00 | ,12 | ,00 |
| | Costa Blanca | 187 | ,03 | ,01 | ,00 | ,29 | ,00 |
| | Costa Brava | 138 | ,02 | ,01 | ,00 | ,15 | ,00 |
| | Costa de Barcelona | 413 | ,08 | ,06 | ,00 | ,53 | ,00 |
| | Costa de Castellón | 31 | ,02 | ,01 | ,00 | ,09 | ,00 |
| | Costa de la Luz | 80 | ,05 | ,02 | ,00 | ,36 | ,00 |
| | Costa de Murcia | 61 | ,12 | ,07 | ,00 | ,49 | ,00 |
| | Costa de Valencia | 36 | ,07 | ,04 | ,00 | ,35 | ,00 |
| | Costa del Sol | 233 | ,04 | ,02 | ,00 | ,27 | ,00 |
| | Costa Dorada | 123 | ,03 | ,01 | ,00 | ,32 | ,00 |
| | Costa Tropical | 18 | ,03 | ,03 | ,00 | ,09 | ,00 |
| | Islas Baleares | 385 | ,02 | ,00 | ,00 | ,28 | ,00 |
| | Islas Canarias | 337 | ,01 | ,00 | ,00 | ,08 | ,00 |

Also very low in coastal locations is the percentage of customers who travel for "Business" being the maximum of 12% on Costa de Murcia

The following chart provides the most information about the data:
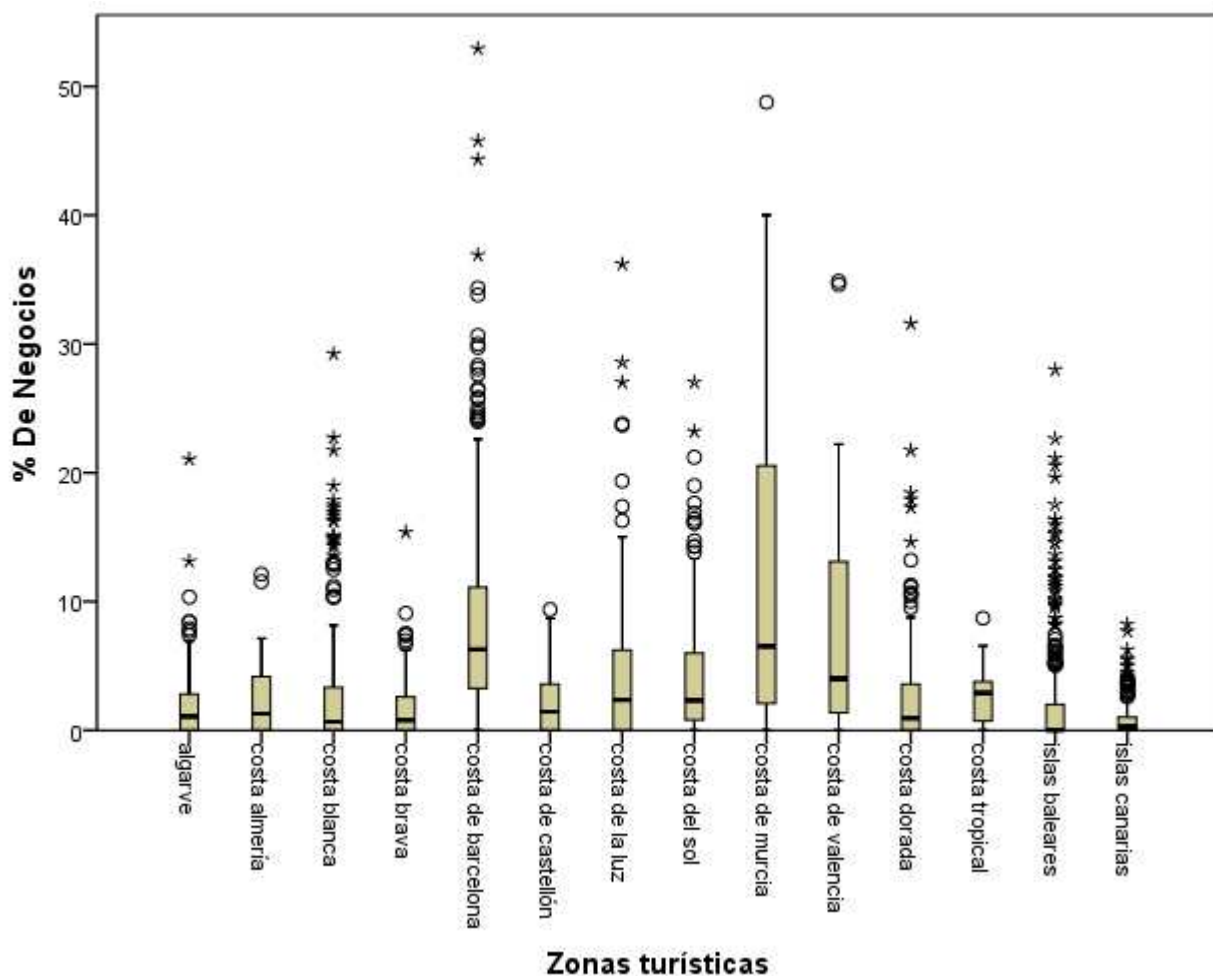
Figure 26: Percentage of comments written by tourists traveling for "Business" in touristic areas

In some locations (Canary Islands, Costa Tropical and Costa de Castellón) there is no hotel that has more than 10% of customers travelling "In Business" (from Customers who write reviews on TripAdvisor)

### 6.2.11 According to the mean percentage of customers travelling with "Friends"

The following table shows the average values ?of the percentage of comments made by customers traveling "With Friends"

Table 41: Percentage of comments "With Friends" for tourist areas

|  |  | [with friends %] | | | | | |
|---|---|---|---|---|---|---|---|
|  |  | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| destino1 | Algarve | 147 | ,09 | ,06 | ,00 | ,49 | ,00 |
|  | Costa de Almería | 22 | ,13 | ,09 | ,02 | ,50 | ,02 |
|  | Costa Blanca | 187 | ,12 | ,11 | ,00 | ,57 | ,00 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Costa Brava | 138 | ,16 | ,13 | ,00 | ,50 | ,00 |
| Costa de Barcelona | 413 | ,12 | ,13 | ,00 | ,59 | ,00 |
| Costa de Castellón | 31 | ,08 | ,07 | ,00 | ,35 | ,00 |
| Costa de la Luz | 80 | ,10 | ,09 | ,00 | ,33 | ,00 |
| Costa de Murcia | 61 | ,09 | ,08 | ,00 | ,37 | ,00 |
| Costa de Valencia | 36 | ,17 | ,16 | ,08 | ,38 | ,03 |
| Costa del Sol | 233 | ,11 | ,10 | ,00 | ,67 | ,00 |
| Costa Dorada | 123 | ,08 | ,07 | ,00 | ,54 | ,00 |
| Costa Tropical | 18 | ,09 | ,08 | ,02 | ,22 | ,02 |
| Islas Baleares | 385 | ,15 | ,10 | ,00 | ,92 | ,00 |
| Islas Canarias | 337 | ,08 | ,06 | ,00 | ,59 | ,00 |

We can see how the hotels on the Costa Brava are those who on average have a higher percentage of customers "Friends" (16%) while the Costa Castellón, Costa Dorada and the Canary Islands are the destinations with the lowest proportion of customers of this type (8%)

The following chart provides the most information about the data:
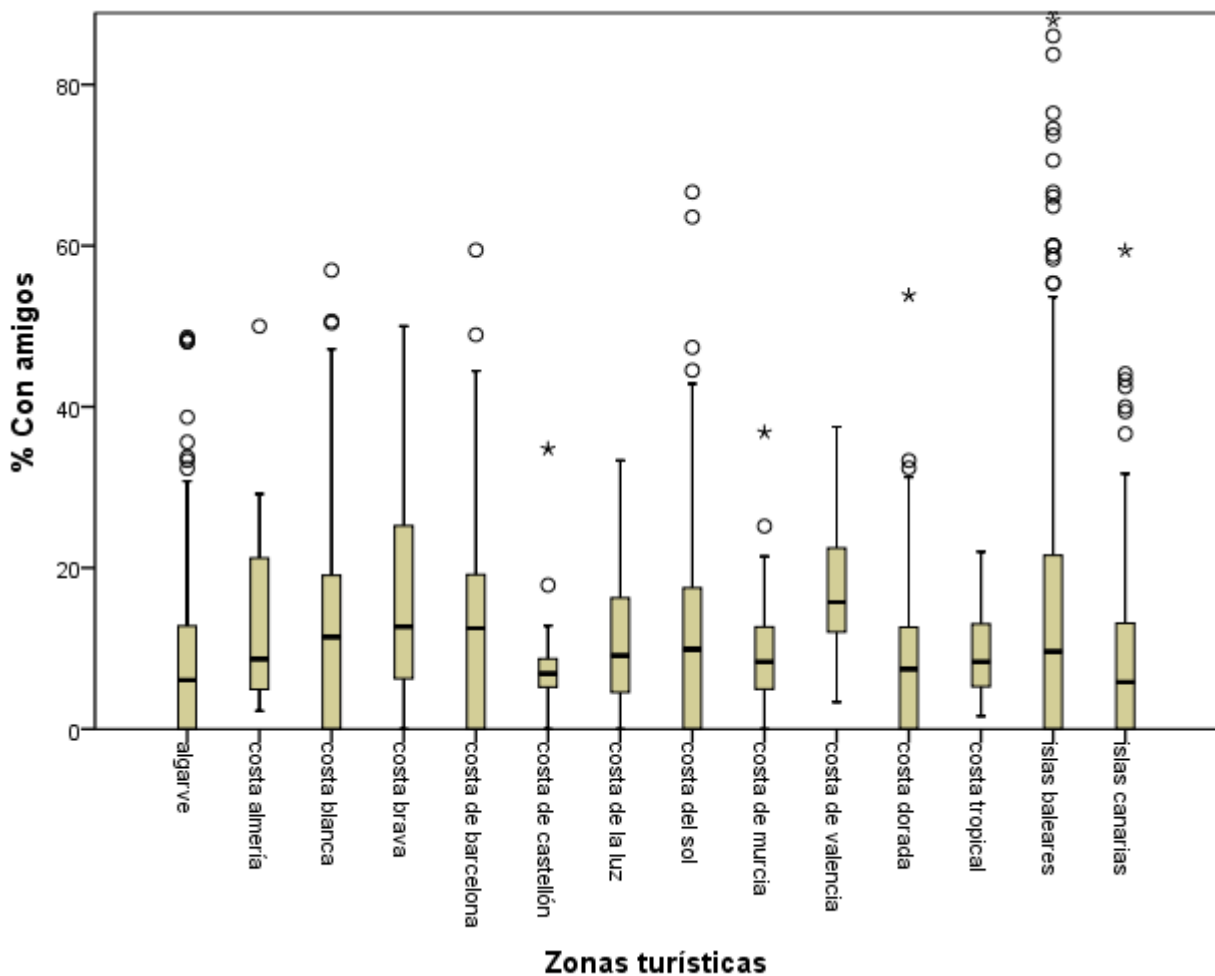


Figure 27: Percentage of comments written by tourists traveling "With Friends" for touristic areas

---

Although the Balearic Islands have hotels where the vast majority of customers (92%) travel with friends

## 6.3 ANALYSIS OF TOURISTIC HOTELS IN COASTAL POINTS OF THE COSTA DEL SOL

For this study we will consider only those points within coastal resorts of the Costa del Sol with at least 10 hotel establishments with more than 20 comments.

### 6.3.1 According to the mean score of the hotels

We see next global figures and mean scores of hotels in the destinations of the Costa del Sol.

Table 42: Overall Rating of tourist spots in hotels

| | | Rating | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recount | Average | Median | Range | Máximum | Mínimum |
| Costa del Sol ( Puntos turísticos ) | Benalmádena | 34 | 3,50 | 3,50 | 3,50 | 5,00 | 1,50 |
| | Fuengirola | 20 | 3,65 | 4,00 | 2,00 | 4,50 | 2,50 |
| | Málaga | 48 | 3,84 | 4,00 | 2,50 | 4,50 | 2,00 |
| | Marbella | 48 | 3,92 | 4,00 | 2,50 | 5,00 | 2,50 |
| | Mijas | 10 | 3,85 | 3,75 | 1,00 | 4,50 | 3,50 |
| | Nerja | 19 | 3,84 | 4,00 | 2,00 | 4,50 | 2,50 |
| | Torremolinos | 51 | 3,62 | 4,00 | 3,50 | 5,00 | 1,50 |

It is remarkable how Benalmádena is far below the rest in terms of scoring their hotels. Also noteworthy is the fact that in Mijas hotels have a score equal to or greater than 3,5. In this regard there is a remarkable dispersion of scores in Benalmadena and Torremolinos hotels with the highest scores (5) and lowest in the area (1.5)

The following figure illustrates these results:
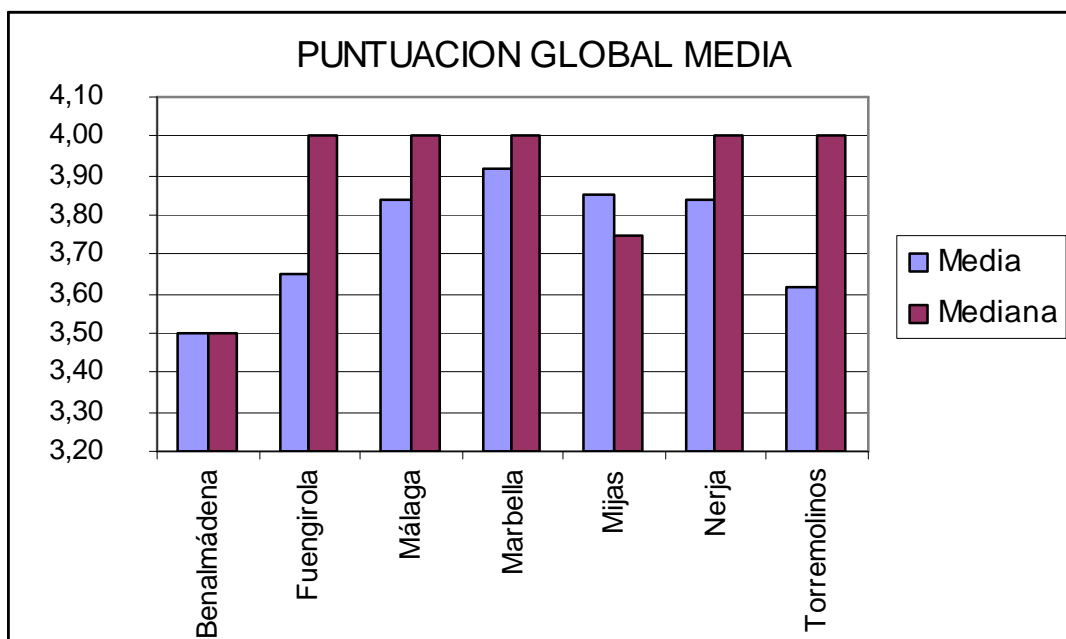
Figure 28: Average overall ranking of hotels on the Costa del Sol by Destination
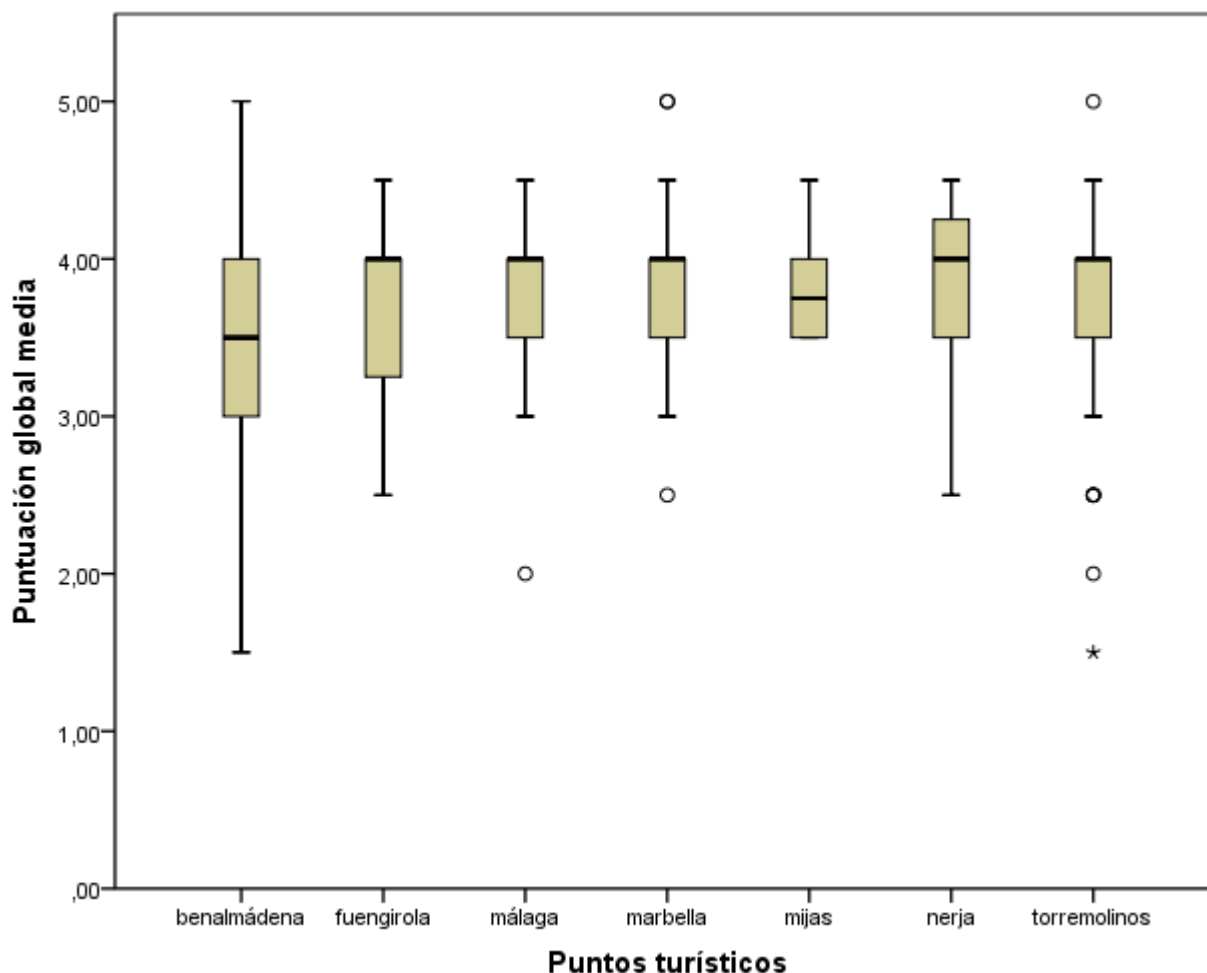


Figure 29: Average overall ranking of hotels on the Costa del Sol by Destination (Box Plot)

## 6.3.2 According to the average size of hotels (measured by number of rooms)

We analyse below the average size of the hotels in the destinations on the Costa del Sol, as measured by number of rooms.

Table 43: Number of rooms in hotels for tourist spots

| | | Número de habitaciones | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recuento | Media | Mediana | Rango | Máximo | Mínimo |
| Costa del Sol (Puntos turísticos) | Benalmádena | 34 | 171 | 168 | 422 | 436 | 14 |
| | Fuengirola | 20 | 203 | 182 | 383 | 408 | 25 |
| | Málaga | 48 | 79 | 59 | 212 | 221 | 9 |
| | Marbella | 48 | 148 | 139 | 395 | 400 | 5 |
| | Mijas | 10 | 137 | 135 | 278 | 302 | 24 |
| | Nerja | 19 | 93 | 58 | 310 | 319 | 9 |
| | Torremolinos | 51 | 204 | 164 | 642 | 650 | 8 |

As for the number of rooms in hotels on the Costa del Sol struck by the small size of the hotels in Málaga Capital. Torremolinos is the biggest yet with 204 rooms on average. Of note is the dispersion in this town with hotels from 8 to 650 rooms
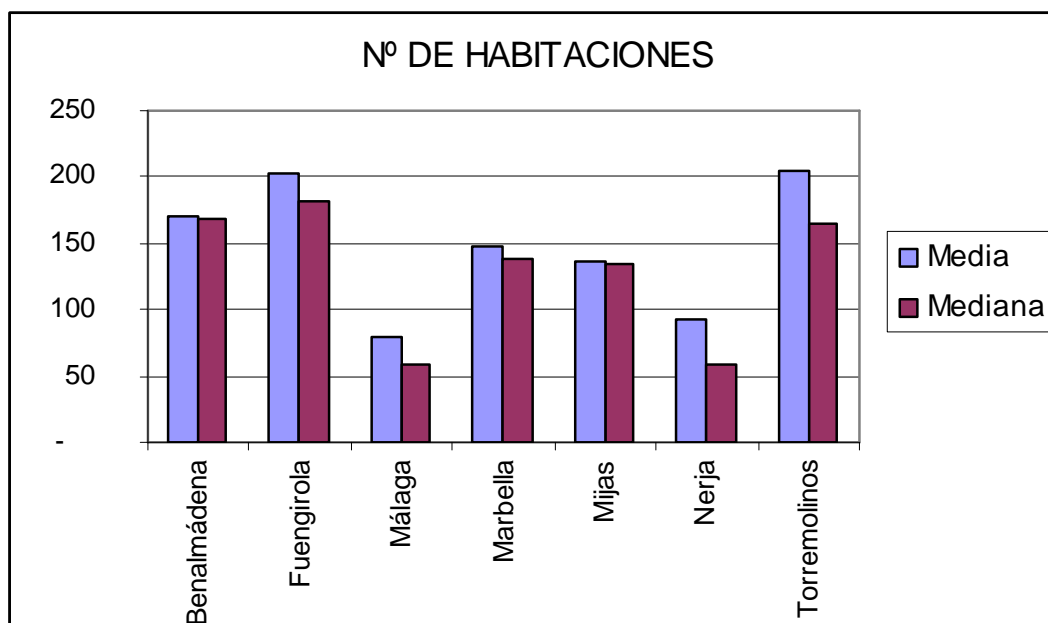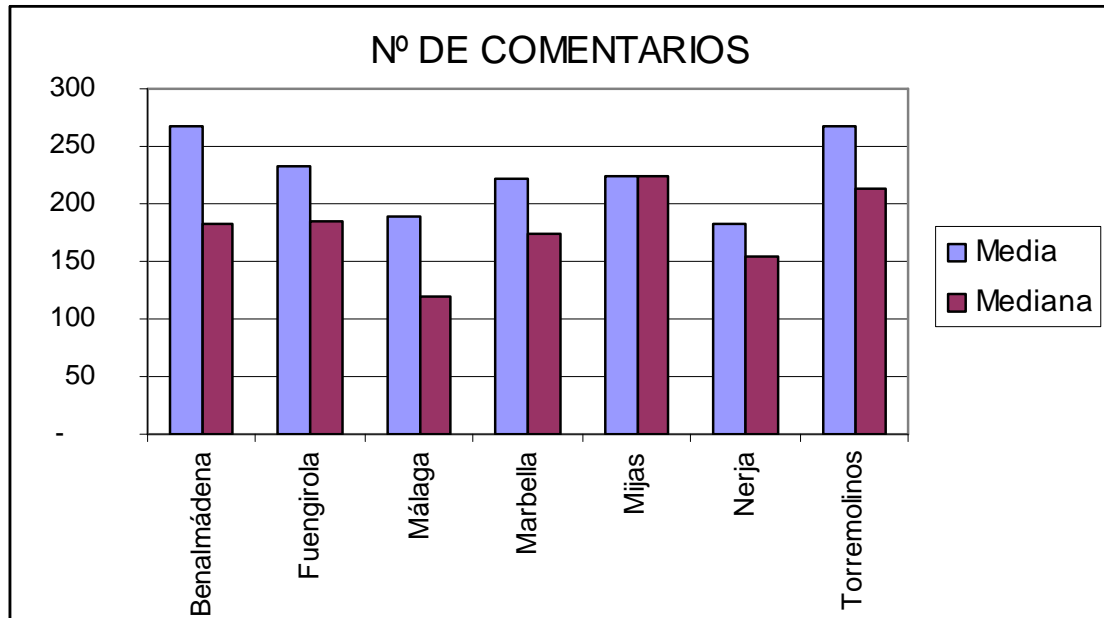


Figure 30: Number of rooms in hotels on the Costa del Sol by Destination
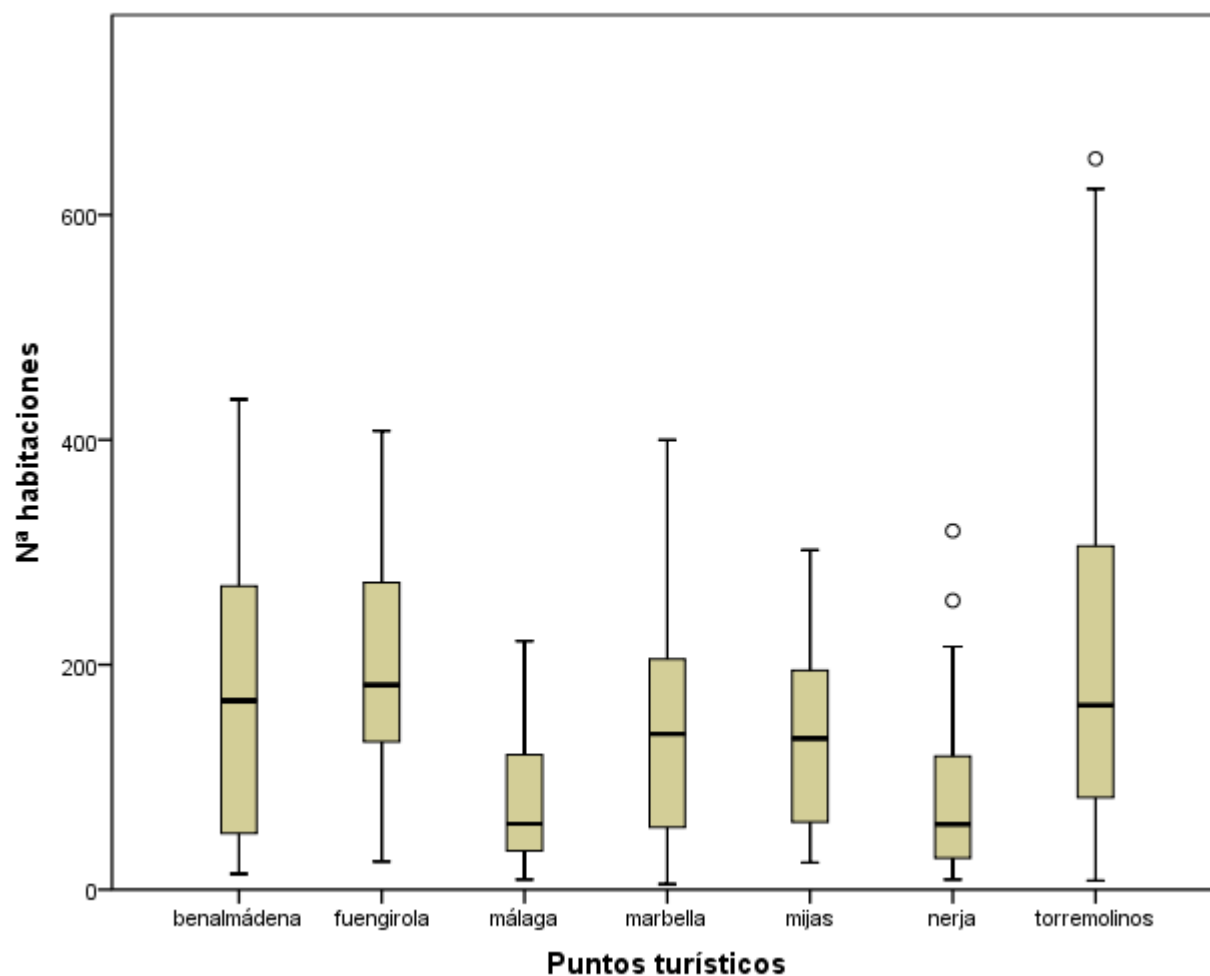
Figure 31: Number of rooms in hotels on the Costa del Sol by Destination (Box Plot)

## 6.3.3 According to the average number of comments about hotels

We turn now to the figures of number of reviews of the hotels in the destinations on the Costa del Sol.

Table 44: Number of comments on hotels for tourist spots

|  |  | Nº de Comentarios | | | | | |
|---|---|---|---|---|---|---|---|
|  |  | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| Costa del Sol ( Puntos turísticos ) | Benalmádena | 34 | 268 | 182 | 27 | 987 | 27 |
|  | Fuengirola | 20 | 233 | 185 | 179 | 555 | 23 |
|  | Málaga | 48 | 190 | 119 | 23 | 655 | 21 |
|  | Marbella | 48 | 221 | 174 | 22 | 809 | 22 |
|  | Mijas | 10 | 225 | 225 | 85 | 434 | 85 |
|  | Nerja | 19 | 182 | 155 | 39 | 521 | 20 |
|  | Torremolinos | 51 | 267 | 212 | 29 | 973 | 20 |

As for the number of reviews, Mijas (the locality with the fewest establishments in the sample) has hotels with more comments, Malaga hotels receive less comments. This variable (number of comments) as we have discussed above provides more information when compared to the number of rooms, which is what we will do in the following paragraph analyzing the values ??for the NCNH ratio



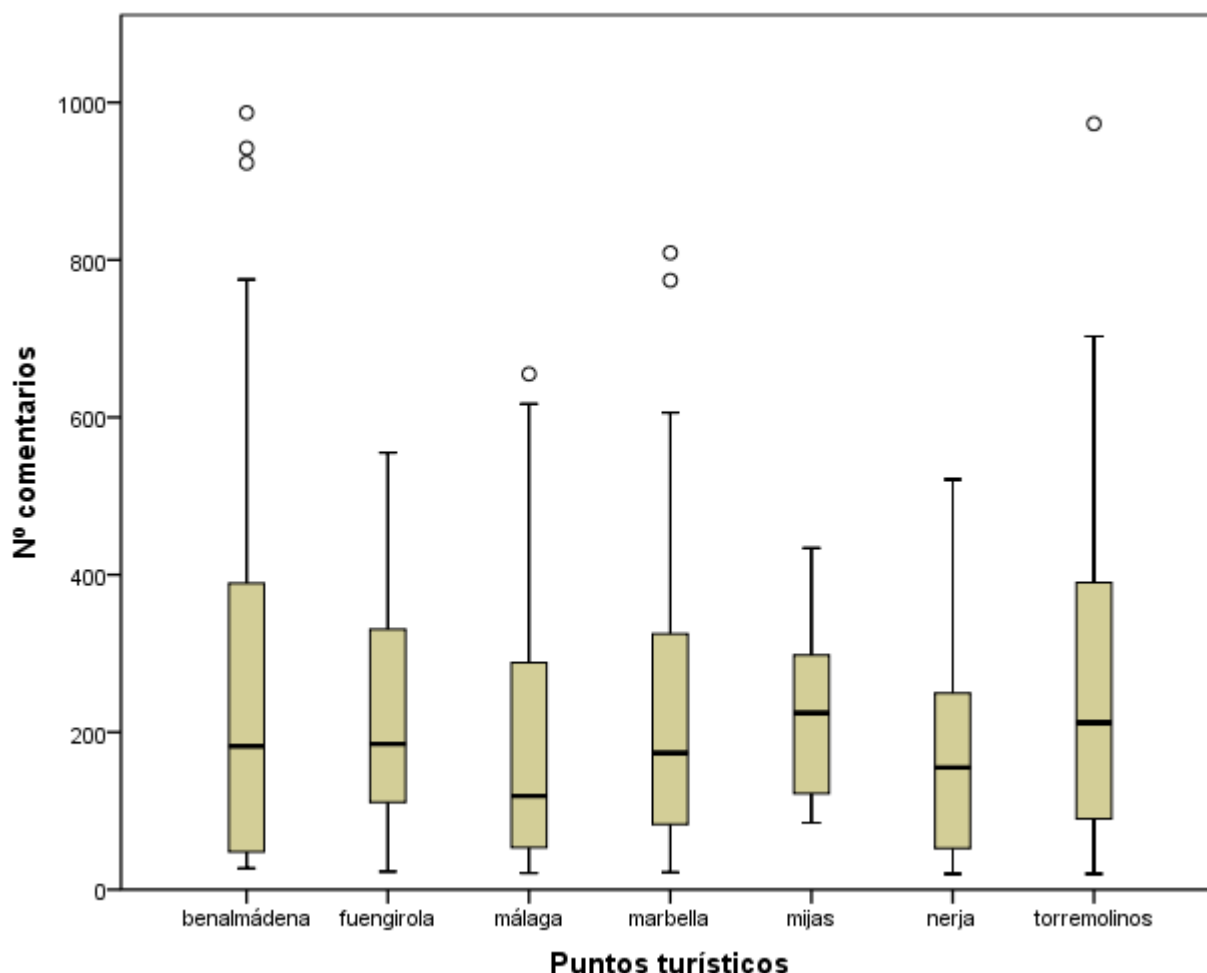Figure 32: Number of comments in hotels on the Costa del Sol by Destination

Figure 33: Number of comments in hotels on the Costa del Sol by Destination (Box diagram)

## 6.3.4 According to the ratio NCNH value (number of comments / number of rooms).

We see next NCNH ratio figures ( Number of comments / Number of rooms) hotels in destinations on the Costa del Sol.

Table 45: NCNH ratio in hotels for tourist spots

| | | Ratio NCNH | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| Costa del Sol ( Puntos turísticos ) | Benalmádena | 34 | 1,71 | 1,27 | 1,71 | 5,74 | 0,19 |
| | Fuengirola | 20 | 1,21 | 1,15 | 0,33 | 2,74 | 0,33 |
| | Málaga | 48 | 2,90 | 2,12 | 0,23 | 16,09 | 0,23 |
| | Marbella | 48 | 3,30 | 1,47 | 0,15 | 33,60 | 0,15 |
| | Mijas | 10 | 2,11 | 1,59 | 0,63 | 5,42 | 0,63 |
| | Nerja | 19 | 2,92 | 2,15 | 0,09 | 9,24 | 0,09 |
| | Torremolinos | 51 | 1,70 | 1,34 | 0,22 | 11,78 | 0,22 |

We note that while it was Mijas that received more comments on average per hotel, Malaga and Nerja are those having higher values ??in the NCNH ratio. Then these destinations which have more comments (proportionally to their size). Hotels in Malaga and Nerja have more than 2 comments per room, are therefore targets for the most active customers.
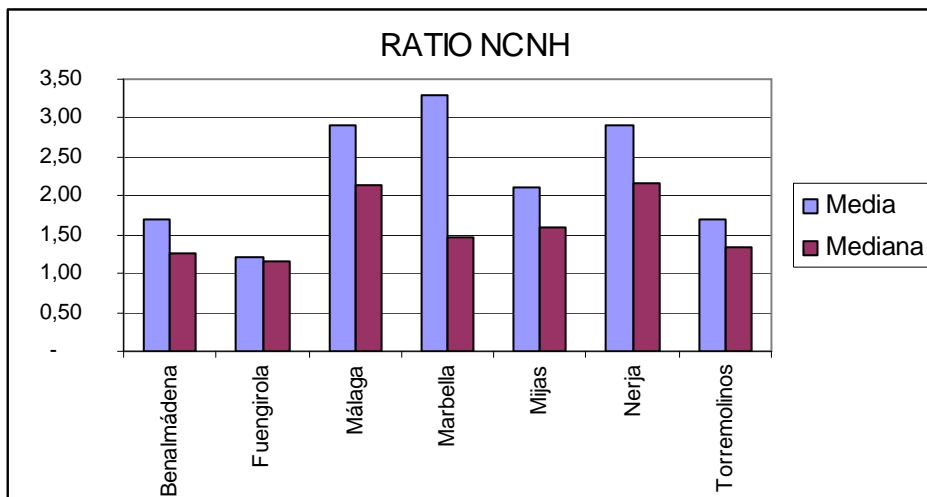


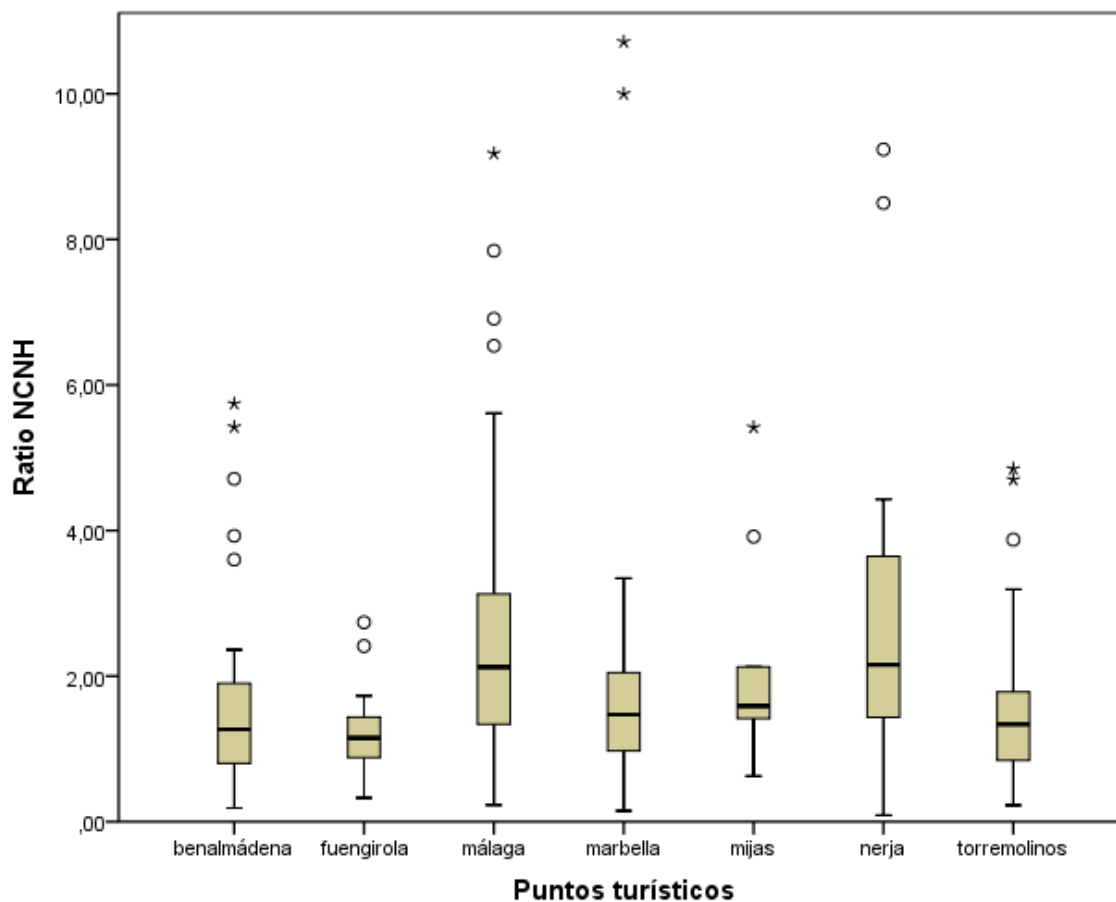Figure 34: NCNH Ratio in hotels on the Costa del Sol by Destination



Figure 35: NCNH Ratio in hotels on the Costa del Sol by Destination

## 6.3.5 According to the mean percentage of "Excellent" and "Very Good" comments

Then the percentage of "Excellent" and "Very good" comments (of total comments) for hotels in the destinations of the Costa del Sol were analyzed.

Table 46: Percentage of "Excellent" and "Very Good" comments in hotels for tourist spots

| | | Porcentaje de Comentarios Excelente + Muy bueno | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| Costa del Sol (Puntos turísticos) | Benalmádena | 34 | 59,03% | 59,26% | 0,00% | 98,51% | 0,00% |
| | Fuengirola | 20 | 62,60% | 67,31% | 30,43% | 87,30% | 30,43% |
| | Málaga | 48 | 72,17% | 74,72% | 15,63% | 94,29% | 15,63% |
| | Marbella | 48 | 73,81% | 77,45% | 52,63% | 98,81% | 24,69% |
| | Mijas | 10 | 73,37% | 73,65% | 52,69% | 86,17% | 52,69% |
| | Nerja | 19 | 68,60% | 75,00% | 75,00% | 94,12% | 23,08% |
| | Torremolinos | 51 | 61,96% | 67,40% | 0,00% | 98,11% | 0,00% |
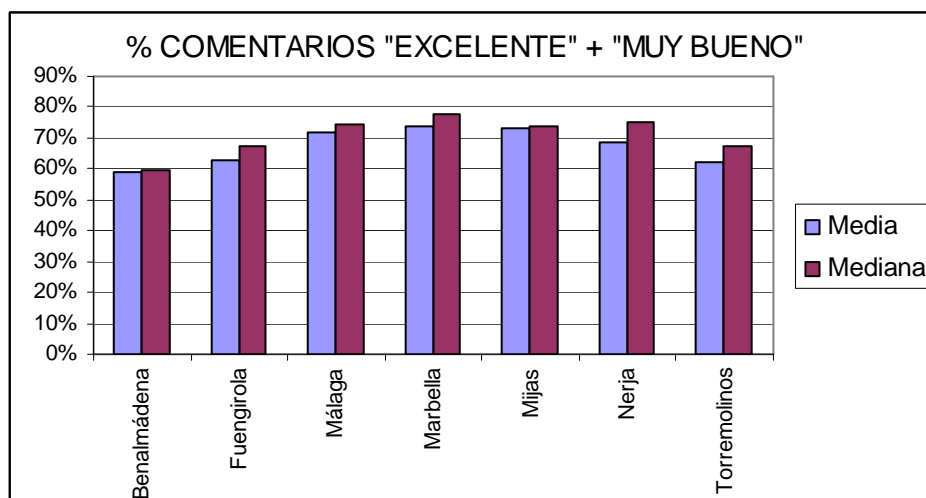


Figure 36: Percentage of "Excellent" and "Very Good" comments in hotels on the Costa del Sol by Destination

Marbella hotels received the best comments. On average, 77% of guests at hotels in Marbella rate their experience as excellent or very good. By contrast, only 59.26% of guests of hotels in Benalmádena valued their stay so positively.
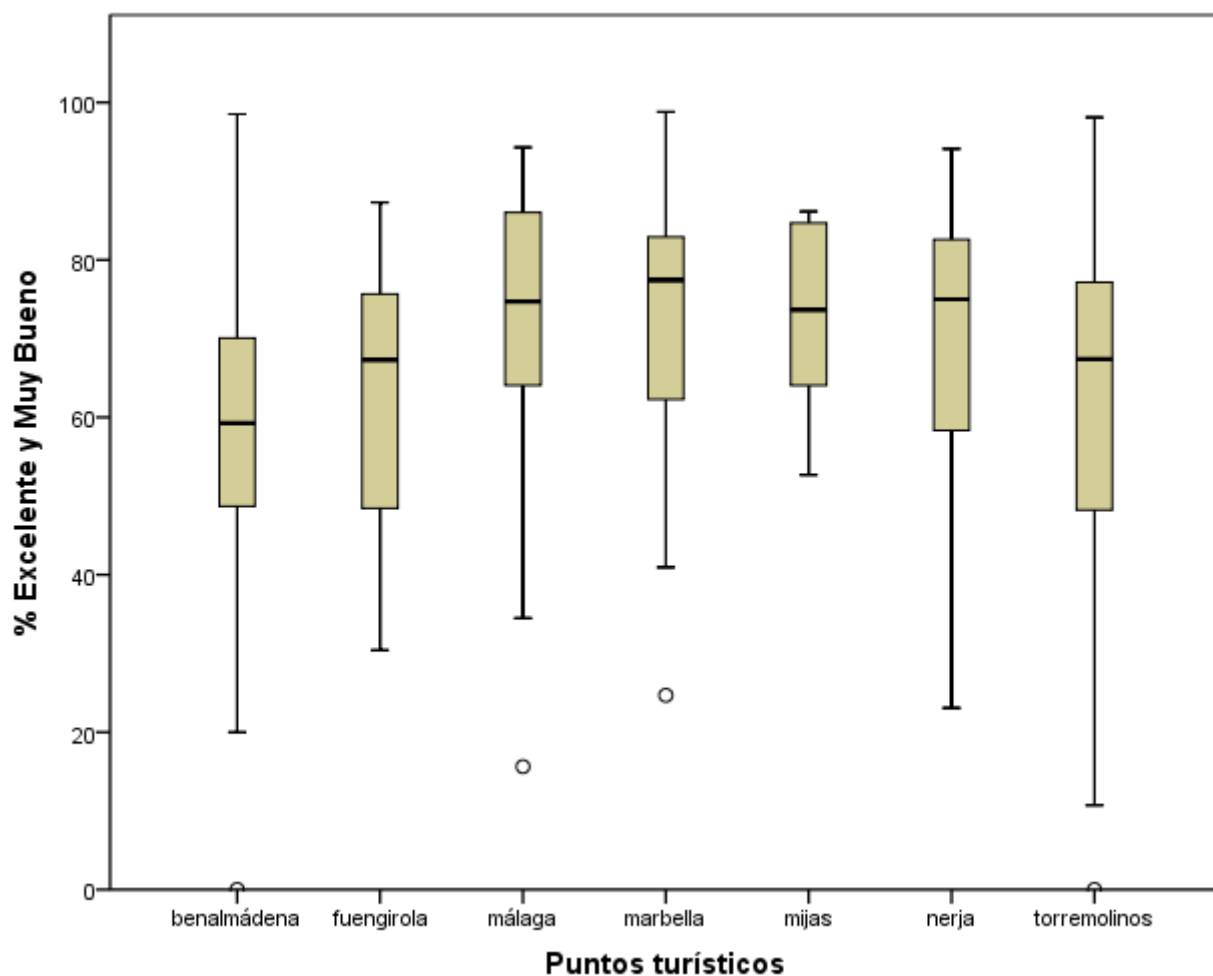
Figure 37: Percentage of "Excellent" and "Very Good" comments in hotels on the Costa del Sol by Destination (Box Plot)

## 6.3.6 According to the mean percentage of "Terrible" and "Bad" comments

Then the percentage of "Terrible" and "Bad" comments (of the totalnumber of comments) about hotels in the destinations on the Costa del Sol were analyzed.

Table 47 Percentage of pessimistic and bad comments in Costa del Sol

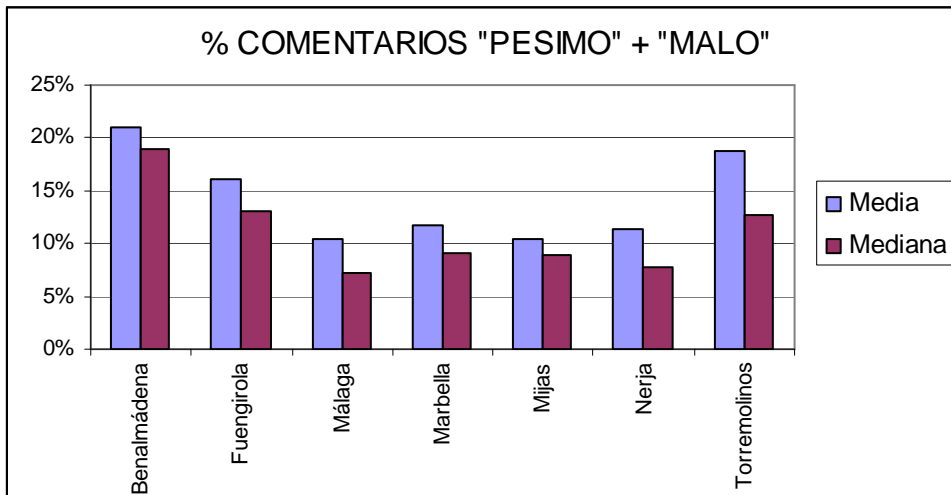| | | Porcentaje de Comentarios Pésimo + Malo | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| Costa del Sol (Puntos turísticos) | Benalmádena | 34 | 21,01% | 18,94% | 25,00% | 90,91% | 0,00% |
| | Fuengirola | 20 | 16,02% | 13,03% | 3,37% | 32,55% | 3,37% |
| | Málaga | 48 | 10,37% | 7,24% | 0,00% | 71,88% | 0,00% |
| | Marbella | 48 | 11,79% | 9,11% | 7,69% | 52,16% | 0,00% |
| | Mijas | 10 | 10,45% | 8,98% | 5,03% | 23,12% | 5,03% |
| | Nerja | 19 | 11,42% | 7,74% | 0,00% | 35,90% | 0,00% |
| | Torremolinos | 51 | 18,81% | 12,63% | 0,00% | 86,96% | 0,00% |

Figure 38: Percentage of "Bad" and "Terrible" comments on hotels on the Costa del Sol
by Destination

Consistent with the previous section, are customers of Benalmadena hotels who (18.94%) value their experiences as poor or very poor. By contrast, only 7.24% of the customers of hotels in Málaga value their experience as negative.
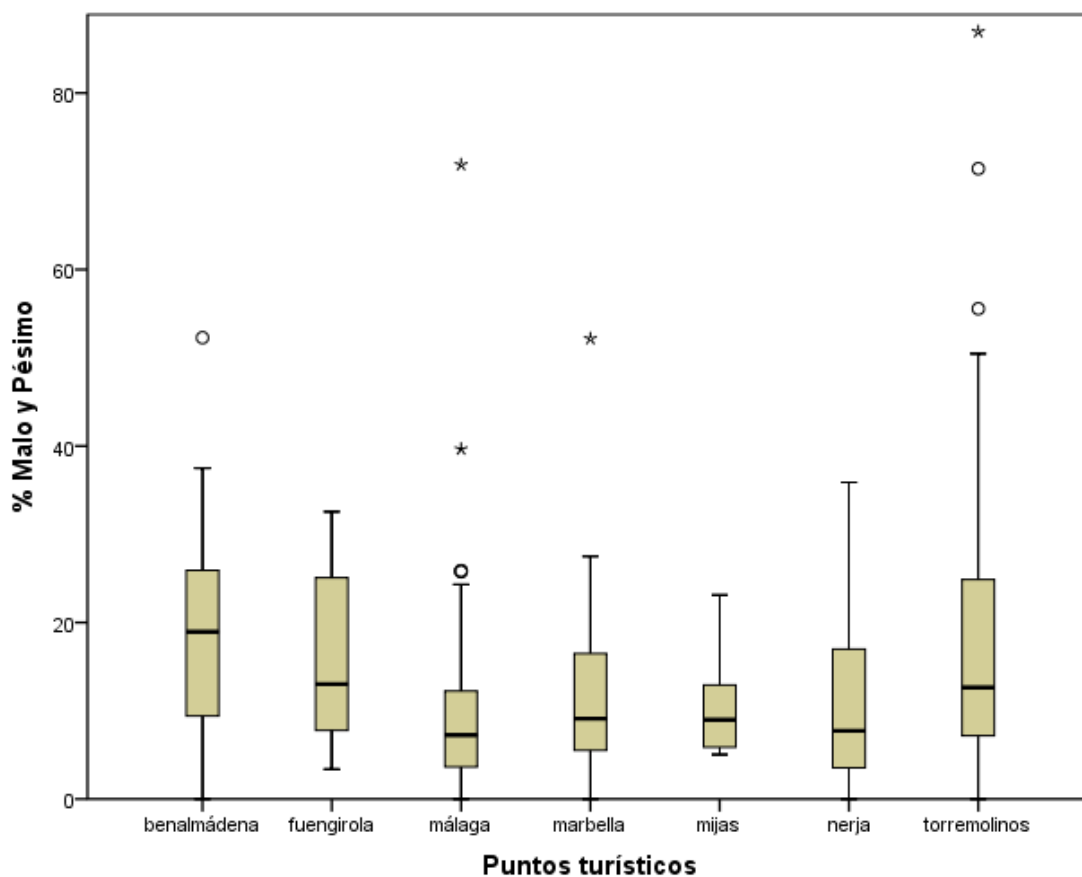


Figure 39: Percentage of "Bad" and "Terrible" comments on hotels on the Costa del Sol
by Destination (Box Plot)

## 6.3.7 According to the mean value of the percentage of "Family" customer comments

We analyze below the average values ??of the percentage of comments made by customers who have traveled as a "Family" (on the number of comments) on the hotels in the destinations on the Costa del Sol.

Table 48: Percentage of "Family" customer comments on hotels for tourist spots

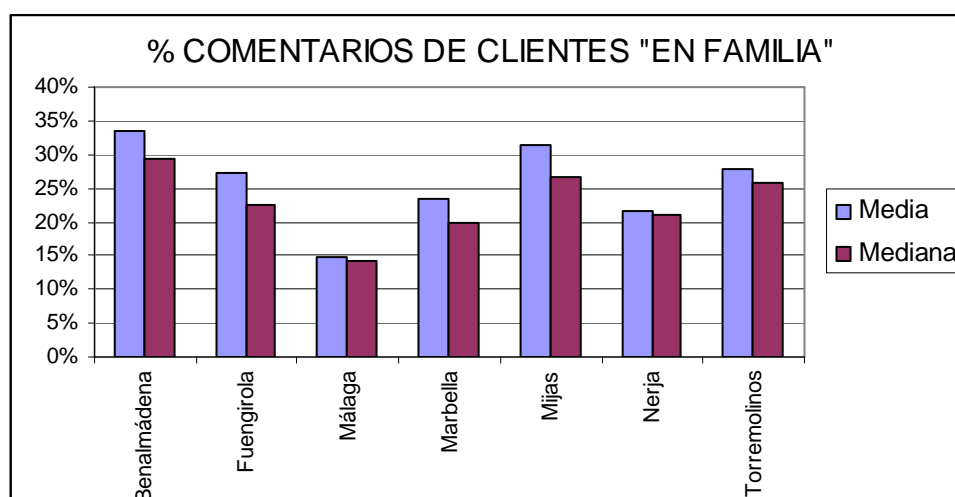| | | % En familia | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| Costa del Sol (Puntos turísticos) | Benalmádena | 34 | 33,48% | 29,27% | 11,11% | 85,87% | 11,11% |
| | Fuengirola | 20 | 27,14% | 22,55% | 5,56% | 71,14% | 5,56% |
| | Málaga | 48 | 14,75% | 14,26% | 11,76% | 41,38% | 3,08% |
| | Marbella | 48 | 23,41% | 20,00% | 15,66% | 74,87% | 5,13% |
| | Mijas | 10 | 31,27% | 26,67% | 8,59% | 55,66% | 8,59% |
| | Nerja | 19 | 21,53% | 21,15% | 8,00% | 45,24% | 8,00% |
| | Torremolinos | 51 | 27,91% | 25,88% | 0,00% | 62,76% | 0,00% |



Figure 40: Percentage of comments from guests traveling in "Family" in hotels in the Costa del Sol by Destination

It is appreciated that Málaga is not considered a family destination by guests commenting about hotels on TripAdvisor and that only 14% of them were identified in this category. By contrast, 30% of customers in other hotels in Benalmadena were identified as "familes".
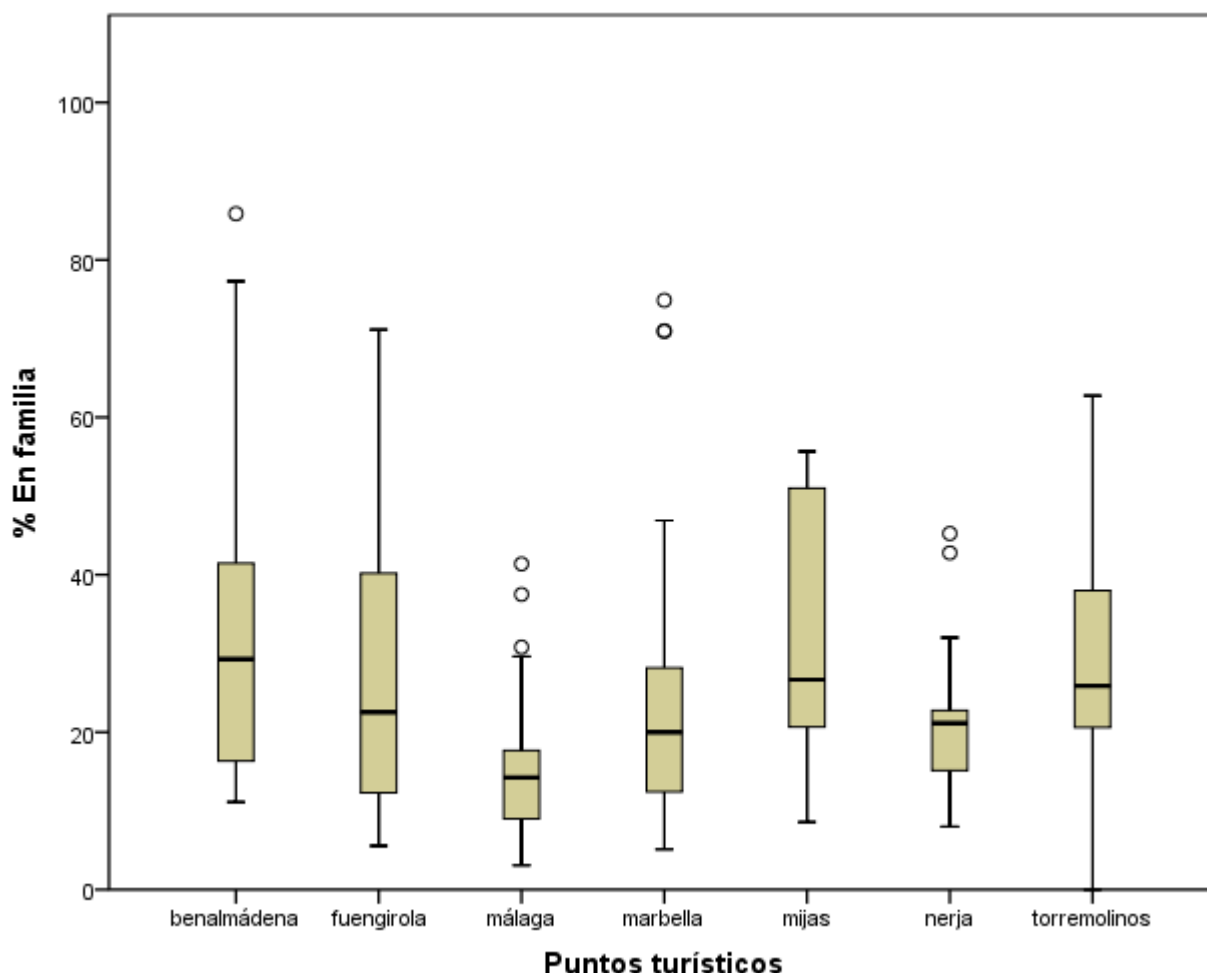
Figure 41: Percentage of comments from guests traveling as "Families" in hotels on the Costa del Sol by Destination (Box Plot)

## 6.3.8 According to the mean value of the percentage of comments made by customers travelling as "Couples" :

We analyze below the average values of the percentage of comments (of total comments) made by customers who have traveled as "Couples" hotels to the destinations on the Costa del Sol.

Table 49: Percentage of "Couples" customer comments in hotels for tourist spots

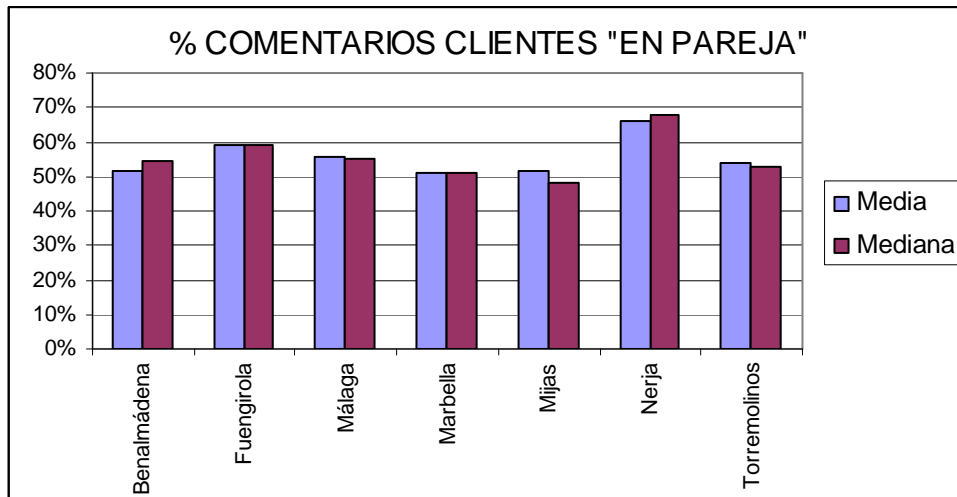| | | % En pareja | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| Costa del Sol ( Puntos turísticos ) | Benalmádena | 34 | 51,34% | 54,72% | 66,67% | 81,03% | 11,96% |
| | Fuengirola | 20 | 59,01% | 59,10% | 18,52% | 84,65% | 18,52% |
| | Málaga | 48 | 55,66% | 55,13% | 35,48% | 74,00% | 35,48% |
| | Marbella | 48 | 51,23% | 51,09% | 16,95% | 83,36% | 16,95% |
| | Mijas | 10 | 51,34% | 48,12% | 24,53% | 80,66% | 24,53% |
| | Nerja | 19 | 65,80% | 68,00% | 38,10% | 81,13% | 38,10% |
| | Torremolinos | 51 | 53,80% | 52,73% | 52,38% | 92,00% | 15,79% |

Figure 42: Percentage of comments from guests traveling as "Couples" in hotels on the Costa del Sol by Destination

Nerja is the destination where a greater number of tourists who write their comments on TripAdvisor identified as "Couples" (68%), and Mijas which has the least number in this section (48.12%)
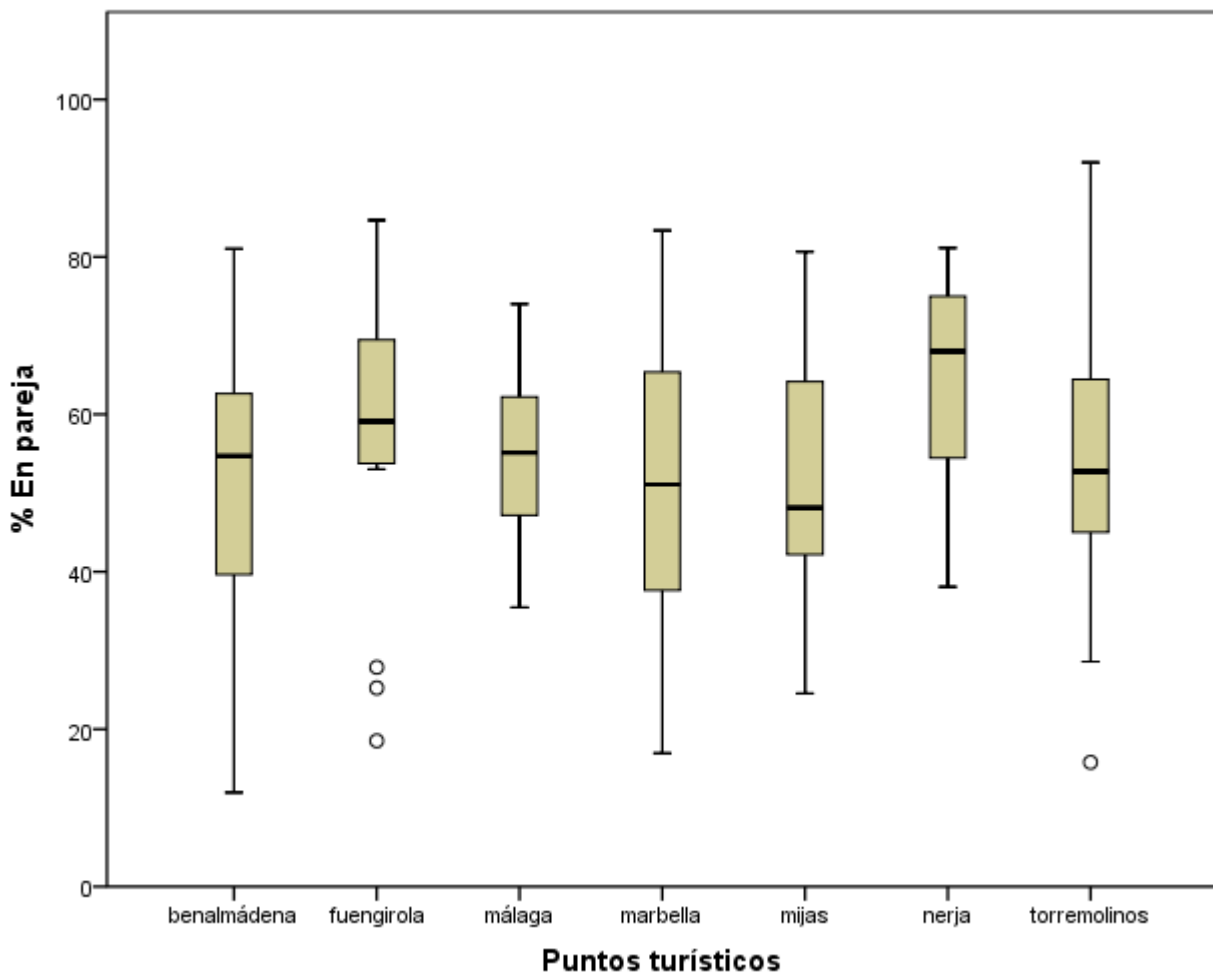


Figure 43: Percentage of comments from guests traveling as "Couples" in hotels on the Costa del Sol by Destination

## 6.3.9 According to the mean value of the percentage of "Single" customer comments

We analyze below the average values ?of the percentage of comments made by customers who have travelled "Single" (of total comments) to hotels in the destinations on the Costa del Sol.

Table 50: Percentage of "single" customer comments in hotels for tourist spots

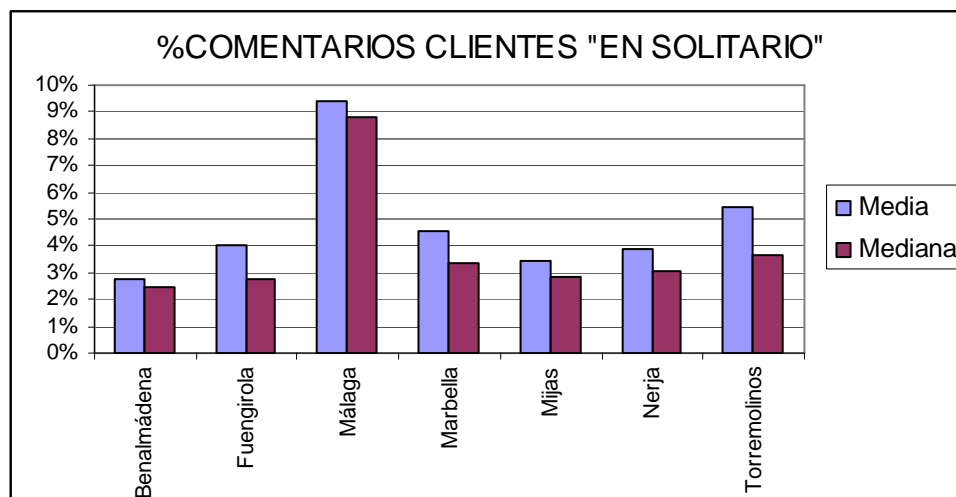| | | % En solitario | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| Costa del Sol ( Puntos turísticos ) | Benalmádena | 34 | 2,73% | 2,44% | 0,00% | 12,24% | 0,00% |
| | Fuengirola | 20 | 4,01% | 2,77% | 0,00% | 16,13% | 0,00% |
| | Málaga | 48 | 9,39% | 8,79% | 0,00% | 27,69% | 0,00% |
| | Marbella | 48 | 4,52% | 3,35% | 0,00% | 20,00% | 0,00% |
| | Mijas | 10 | 3,46% | 2,86% | 1,56% | 7,61% | 1,56% |
| | Nerja | 19 | 3,85% | 3,08% | 0,00% | 10,53% | 0,00% |
| | Torremolinos | 51 | 5,47% | 3,64% | 0,00% | 26,92% | 0,00% |



Figure 44: Percentage of comments from "Single" guests staying at hotels on the Costa del Sol by Destination

Apart from Málaga with 8.79% of its tourists traveling alone "single", other destinations on the Costa del Sol have very low figures in this section
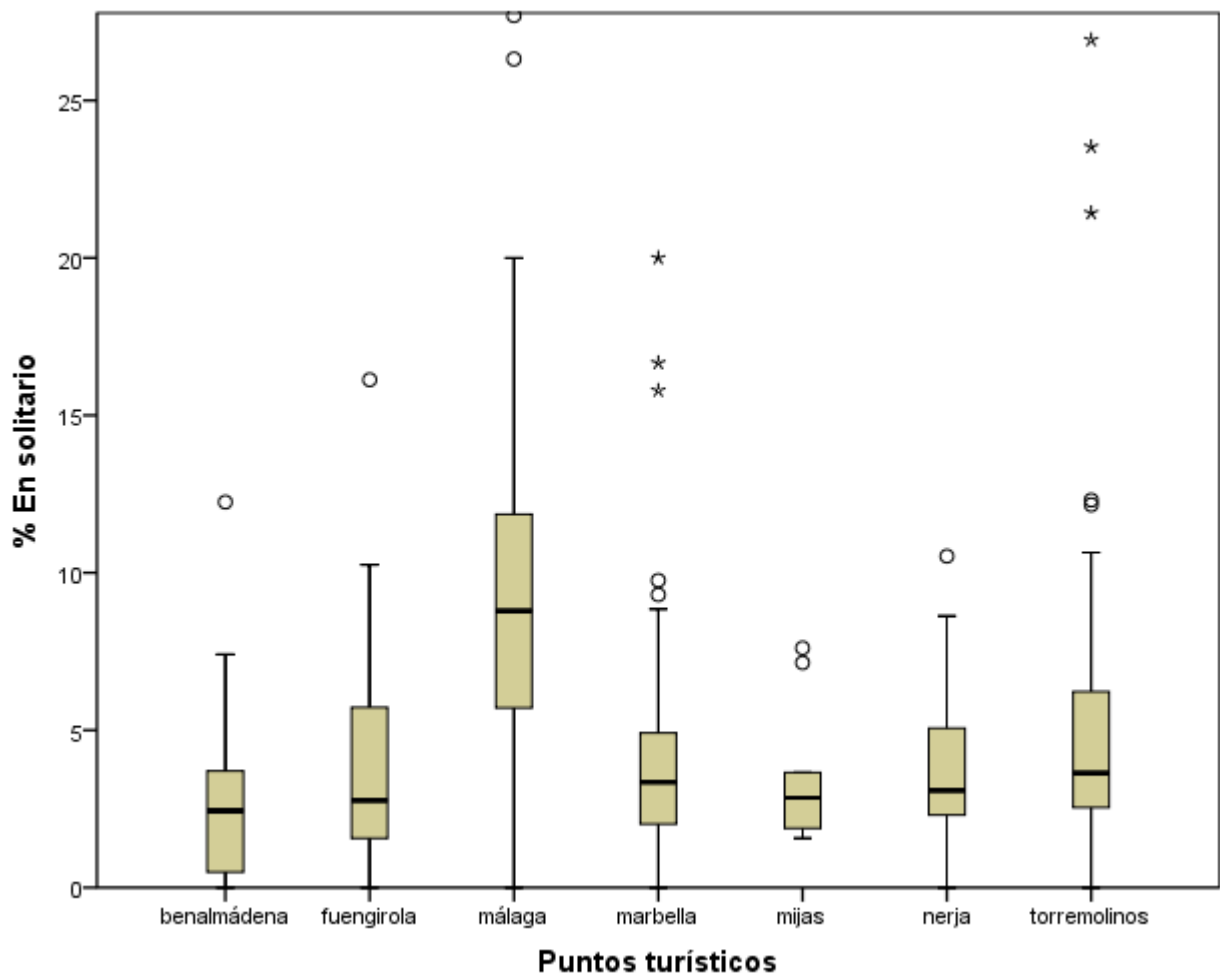
Figure 45: Percentage of comments from guests traveling "Solo" at hotels on the Costa del Sol by Destination (Box Plot)

## 6.3.10 According to the mean value of the percentage of "Business" customer comments

We analyze below the average values ?of the percentage of comments made by customers who traveled for "Business" (of the total number of comments) on the hotels in the destinations on the Costa del Sol.

Table 51: Percentage of "Business" customers´ comments on hotels
for tourist spots

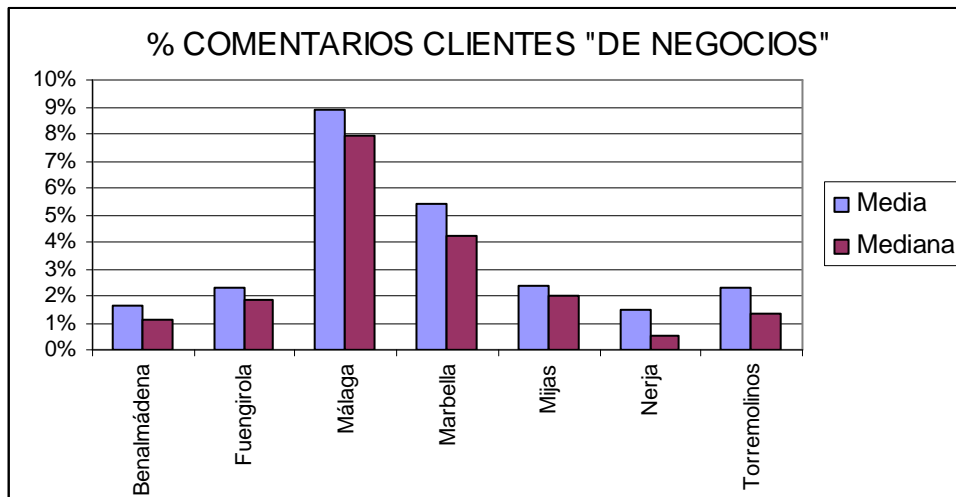| | | % De negocios | | | | | |
|---|---|---|---|---|---|---|---|
| | | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| Costa del Sol ( Puntos turísticos ) | Benalmádena | 34 | 1,62% | 1,12% | 0,00% | 8,82% | 0,00% |
| | Fuengirola | 20 | 2,28% | 1,85% | 0,00% | 6,45% | 0,00% |
| | Málaga | 48 | 8,87% | 7,90% | 0,00% | 27,03% | 0,00% |
| | Marbella | 48 | 5,41% | 4,25% | 0,00% | 21,21% | 0,00% |
| | Mijas | 10 | 2,40% | 1,99% | 1,12% | 4,88% | 1,12% |
| | Nerja | 19 | 1,47% | 0,52% | 0,00% | 8,33% | 0,00% |
| | Torremolinos | 51 | 2,30% | 1,32% | 0,00% | 14,29% | 0,00% |



Figure 46: Percentage of comments from guests traveling on "Business" in hotels on the Costa del Sol by Destination

As in the previous section there is very small percentage of customers who write their reviews on TripAdvisor and who travel to the Costa del Sol for business.
Only Málaga presents a significant figure (7.90%), the rest with the exception of Marbella (4.25%) are less than 2%
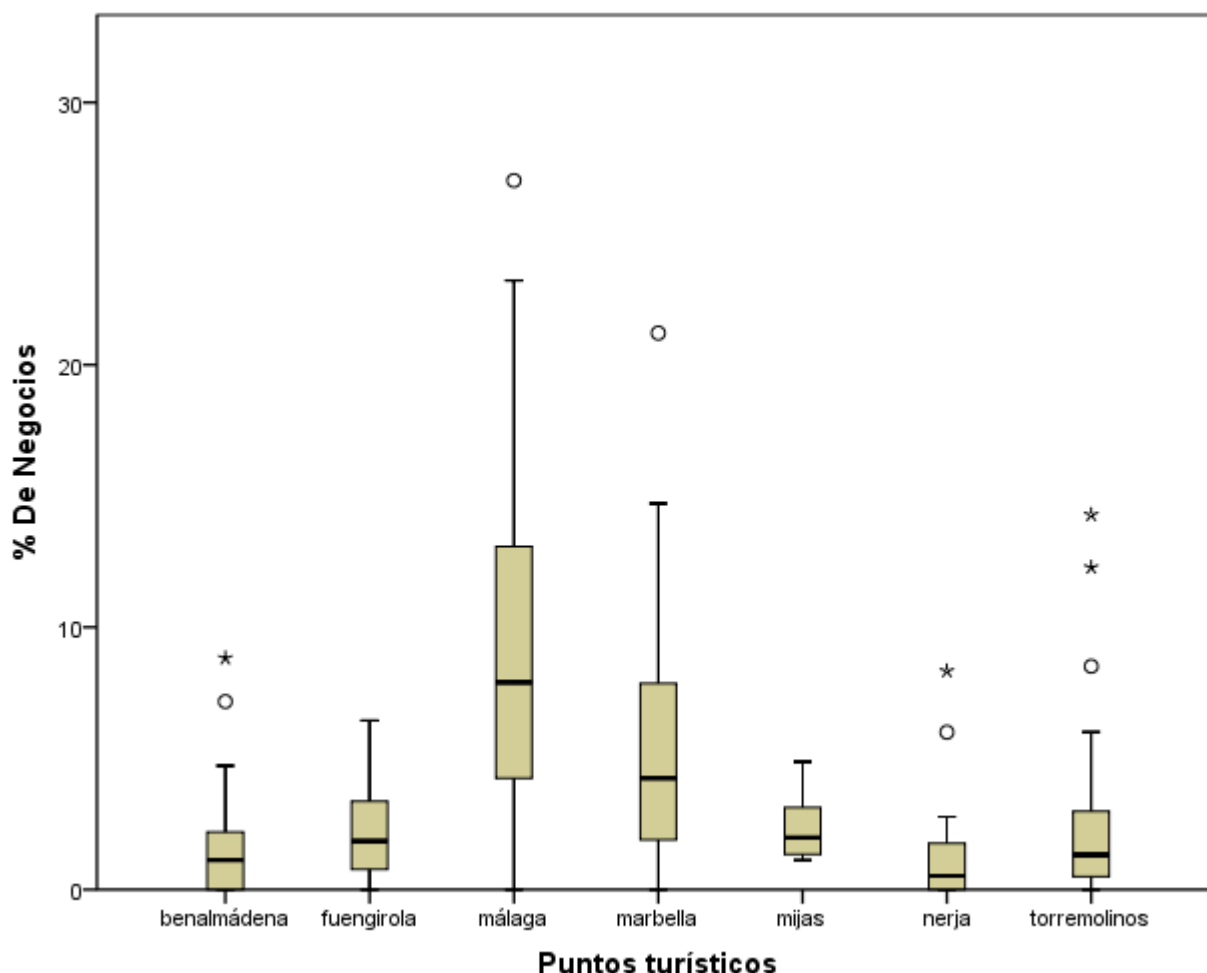
Figure 47: Percentage of comments from guests traveling on "Business" in hotels on the Costa del Sol by Destination

## 6.3.11 According to the mean value of the percentage of customers` comments travelling "With Friends"

We analyze below the average values ?of the percentage of comments made by customers who traveled with "Friends" (of the number of comments) on the hotels in the destinations on the Costa del Sol.

Table 52: Percentage of customers` "With Friends"  comments on hotels for tourist spots

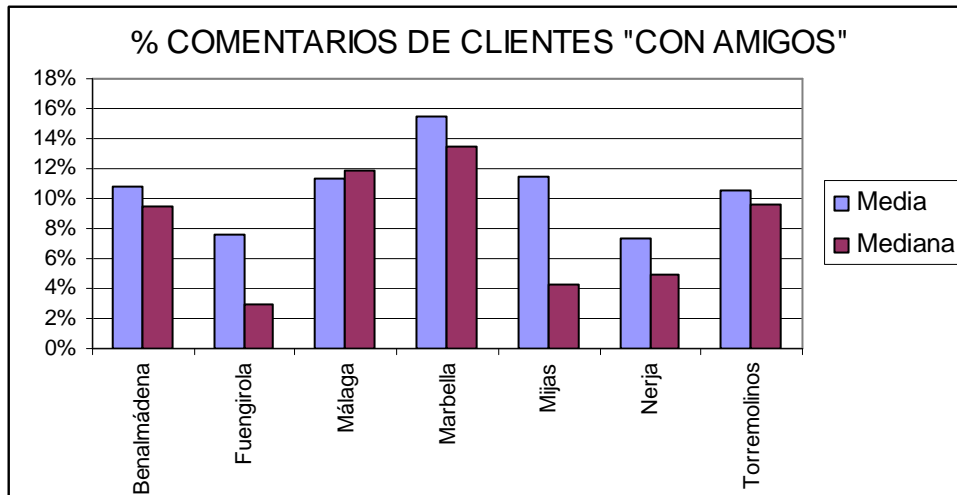|  |  | % Con amigos | | | | | |
|---|---|---|---|---|---|---|---|
|  |  | Recuento | Media | Mediana | Moda | Máximo | Mínimo |
| Costa del Sol ( Puntos turísticos ) | Benalmádena | 34 | 10,84% | 9,42% | 0,00% | 42,86% | 0,00% |
|  | Fuengirola | 20 | 7,56% | 2,94% | 0,00% | 32,04% | 0,00% |
|  | Málaga | 48 | 11,32% | 11,87% | 0,00% | 32,00% | 0,00% |
|  | Marbella | 48 | 15,44% | 13,41% | 0,00% | 66,67% | 0,00% |
|  | Mijas | 10 | 11,53% | 4,27% | 0,00% | 44,53% | 0,00% |
|  | Nerja | 19 | 7,36% | 4,92% | 0,00% | 22,22% | 0,00% |
|  | Torremolinos | 51 | 10,52% | 9,57% | 0,00% | 47,37% | 0,00% |

Figure 48: Percentage of comments from guests traveling as "Friends" in hotels on the Costa del Sol by Destination
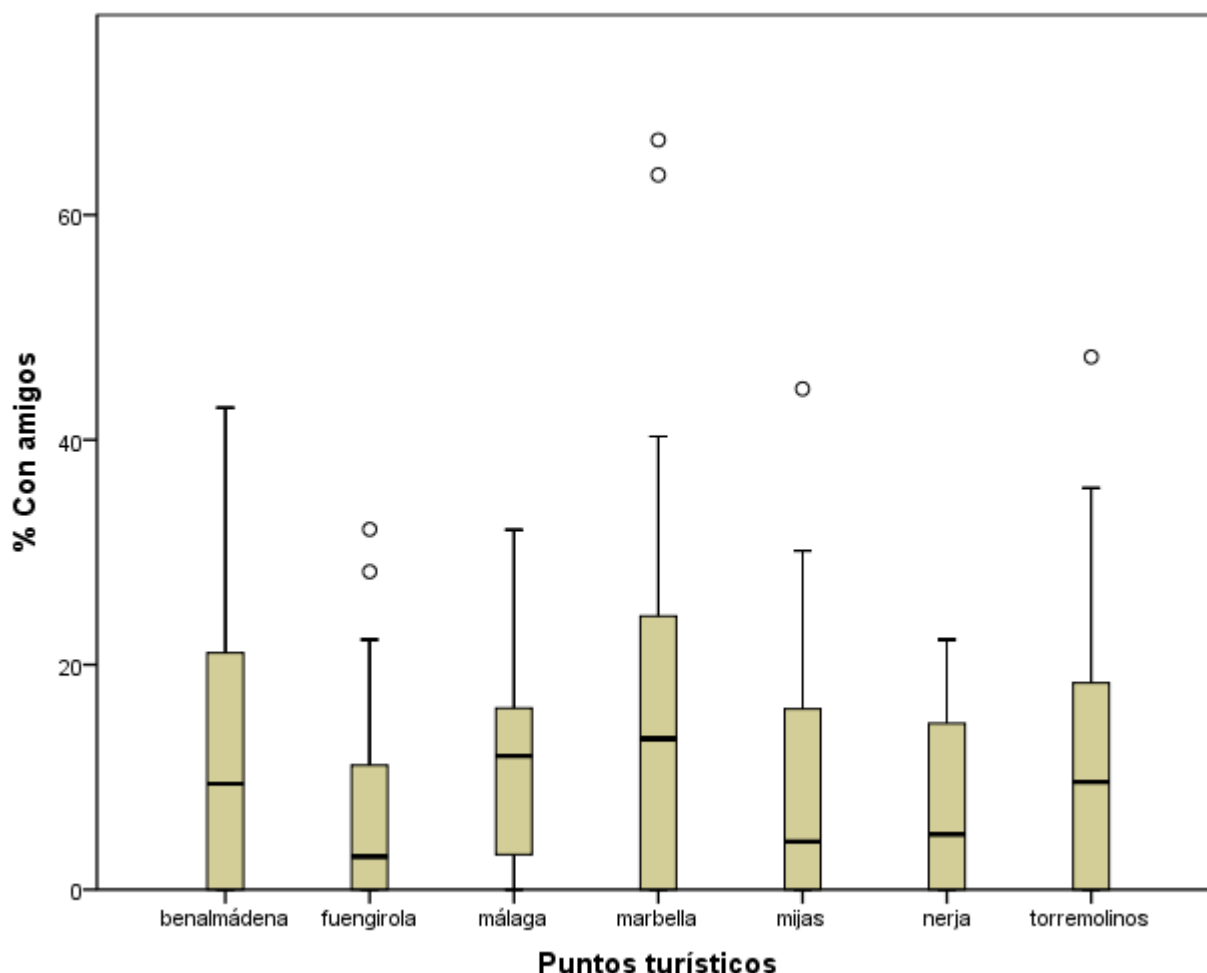
Figure 49: Percentage of comments from guests traveling as "Friends" in hotels on the Costa del Sol by Destination

Marbella is the destination where most tourists travel with "Friends". On the contrary, Fuengirola only receives 2.94% of this type of tourists.

# 7 CONCLUSIONS

TripAdvisor's success is not only in the number of visitors to its site (more than 200 million new visitors per month) or the number of reviews that are written on the site (one per second), but above all in its enormous capacity to influence the planning process and decisions of tourists, having replaced the traditional word of mouth (WOM) with a new way of sharing experiences (e-WOM), and have been able to generate what some authors have called "The Index Effect" which is nothing but the improvement in the quality of service provided by hotels due to the risk of receiving bad reviews from their customers on this website.

Although it should be considered an intermediary, some authors have stressed the responsibility acquired by TripAdvisor as a main figure in the tourism sector.
Given its great influence already mentioned and the way that it controls the false comments posted on its pages which can damage the image of a free listing, or the way that it classifies hotels, which can lead to increased sales for establishments conveniently situated in the ranking if they

are, should be scrupulously analyzed by researchers and students of tourism, otherwise we would not be able to talk about them as figures who have contributed to the sector with innumerable dynamism and other positive elements (such as transparency, new ways of engaging customers, help in decision making, improvement in hotel management ...) but on the contrary, we would be in the presence of an organization with the ability to manipulate and thus corrupt the free market, by action or by omission of responsibility.

There is still widespread opinion among experts and practitioners that TripAdvisor has changed the ways of doing things in the industry, both by the tourist information available when planning your trip and as part of hoteliers how to manage relationships with customers, we must be aware that the Index ranking is a tool that may determine the future viability of a facility and therefore in-depth analysis of how our algorithm works for hotels is required. And one must check if our examples are the norm or a curious exception.

Our work has shed light on some aspects of the operation of International coastal destinations in Spain and Portugal.

1-Firstly, we can highlight the most significant fact, which is that most tourists favorably rate their hotel experiences.

This calls into question some common views on the sector:

(i) The "fear" of TripAdvisor professionals, and customer feedback. As our data shows, this fear is unfounded as most of the reviews are very favorable. The 52.56% of the reviews are high (4 - 4,5 - 5) compared to only 4.98% of low scores (1,5 - 2 - 2,5). The argument that those who post comments are usually dissatisfied customers does not seem to be unfounded.
If we understand that a disgruntled customer is especially motivated to write and share their bad experience and, as we have seen, there are only 4.98% of low scores, it is reasonable to think that the quality of hotel service in our country is more than acceptable .

(ii) The idea that we only think about very good or very bad experiences fits the reality of our numbers. A widespread view in the tourism industry is that comments and ratings contained only positive and negative extremes. The data suggests that this is not so. There are hardly any extreme positive or negative cases, extreme meaning one (1) or (5). Only 0.9% of positive extreme cases (5) and even less, 0.18% negative extreme cases (1.5) (remember that there are really extreme negative cases (1). Additionally there is a very large number (42.47%) of "normal" Ratings (3 - 3,5)

2-Second hotel size measured by the number of rooms has a significant relationship with the mean score of the hotel. There is some evidence to allow us to indicate an inverse relationship between the size of the hotel and the mean score. That is, by decreasing the size of the hotel (as measured by number of rooms) we averaged with higher scores.

3-The hotels should encourage their clients to discuss their experiences and values. Firstly, because we have seen the trend to be mostly positive, but also their number has some influence on the mean score of the hotel. However the number of comments received by a hotel is not as important as the NCNH ratio (# of comments / number of rooms). As a hotel gets more comments and the ratio increases on average with better scores.

4-We can appreciate a significant relationship between the number of hotel rooms and the NCNH value. On average, the higher value, the lower the hotel ratio is.

These results have important practical belief implications for the sector, which should strive to encourage their active participation in customers´ social networks and especially on sites like Tripadvisor with the contribution of those who freely share their views and opinions, help future

tourists to make their decisions more effectively, reduce risks to their experiences more satisfactorily and can have their expectations met.

These platforms also help hoteliers´ decision-making and the ability to quickly detect and correct mistakes or analyze the competition. We further believe that Tripadvisor is ideal for the promotion of small and medium-sized independent hotels, as the results of our research indicate, small hotels get better scores on average than larger ones, allowing them to be placed in good positions for gaining visibility ranking against the big chains. Therefore TripAdvisor becomes the best showcase for this segment, otherwise, it could hardly compete in advertising budgets, media or online search engine positioning with the "giants" of the sector.

The heads of businesses in touristic destinations can also take advantage of the information offered by these platforms to better understand the profile of the tourists, what they like or what they dislike, and can engage in conversations with those who are planning their trip, not only to attract them but also with those who have already visited destinations who can provide invaluable information to those who could well become prescribers of it.

Finally for researchers into tourism, there are great lines of research around online reviews so that we are in a field full of opportunities and questions waiting to be answered.

# BIBLIOGRAPHY

Abián, M. Á. (2005). El futuro de la web.

Akiva, N., & Schler, J. (2008). Mining and Visualizing Online Web Content Using BAM: Brand Association Map TM, 170?171.

Bjørkelund, E., Burnett, T. H., & Nørvåg, K. (2012). A study of opinion mining and visualization of hotel reviews. Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services - IIWAS  ?12, 229. doi:10.1145/2428736.2428773

Bronner, F., & De Hoog, R. (2011). Vacationers and eWOM: Who Posts, and Why, Where, and What? Journal of Travel Research, 50(1), 15?26. Retrieved from 10.1177/0047287509355324

Brown, B. (2012). Beyond Recommendations: Local reviews web sites and their impact. ACM Transactions on Computer-Human Interaction, 19(4), 1?24. doi:10.1145/2395131.2395134

Buttinger, C. (2010). Extracting Room Prices from Web Tables - an Ontology-Aware Approach. Information and communication technologies in tourism.

Callarisa-Fiol, L. J., Sánchez-García, J., Moliner Tena, M. Á., & Forgas-Coll, S. (2012). LA IMPORTANCIA DE LAS COMUNIDADES VIRTUALES PARA EL ANALISIS DEL VALOR DE MARCA.EL CASO DE TRIPADVISOR EN HONG KONG Y PARIS. Papers de turisme, (54), 89?115.

Carrasco, R. A., & Carmona, K. (2011). Data Mining:Aplicaciones a marketing en el sector turístico.

Casaló, L. V., Flavián, C., & Guinalíu, M. (2011). Understanding the intention to follow the advice obtained in an online travel community. Computers in Human Behavior, 27(2), 622?633. doi:10.1016/j.chb.2010.04.013

Connor, P. O. (2010). Managing a Hotel ? s Image on TripAdvisor, 754?772. doi:10.1080/19368623.2010.508007

Deutch, D., & Milo, T. (2012). Mob data sourcing. Proceedings of the 2012 international conference on Management of Data - SIGMOD  ?12, 581. doi:10.1145/2213836.2213905

Esuli, A., & Sebastiani, F. (2006). SENTIWORDN ET?: A Publicly Available Lexical Resource for Opinion Mining, 417–422.

García-crespo, Á., López-Cuadrado, J. L., Colomo-palacios, R., González-carrasco, I., & Ruiz-Mezcua, B. (2011). Sem-Fit: A semantic based expert system to provide recommendations in the tourism domain. Expert Systems with Applications, 38(10), 13310?13319. doi:10.1016/j.eswa.2011.04.152

Gerdes Jr, J., & Stringam, B. B. (2008). Addressing researchers? quest for hospitality data:

Mechanism for collecting data from web resources. Tourism Analysis, 13, 309?315.

Ghose, a., Ipeirotis, P. G., & Li, B. (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. Marketing Science, 31(3), 493?520. doi:10.1287/mksc.1110.0700

Gretzel, U.,Hyan-Yoo,K., & Purifoy,M. (2007). Online Travel Review Study. Laboratory for intelligent System in Tourism.

Hu, X., & Wu, B. (2009). Classification and Summarization of Pros and Cons for Customer Reviews. 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, 73?76. doi:10.1109/WI-IAT.2009.234

Jain, D., Juman, D., Quinby, D., & Rauch, M. (2012). Social Media in Travel 2012. Social Networks & Traveler Reviews. PhocusWright.

Johnson, P., Sieber, R., & Magnien, N. (2012). Automated web harvesting to collect and analyse user-generated content for tourism. Current Issues in Tourism, 15(3), 293?299. Retrieved from http://ejournals.ebsco.com/direct.asp?ArticleID=4C9A81AF07F40C914150

Jurca, R., Garcin, F., Talwar, A., & Faltings, B. (2010). Reporting incentives and biases in online review forums. ACM Transactions on the Web, 4(2), 1?27. doi:10.1145/1734200.1734202

Kwok, L., & Yu, B. (2012). Spreading Social Media Messages on Facebook: An Analysis of Restaurant Business-to-Consumer Communications. Cornell Hospitality Quarterly, 54(1), 84?94. doi:10.1177/1938965512458360

Lee, H. ?Andy?, Law, R., & Murphy, J. (2011). Helpful Reviewers in TripAdvisor, an Online Travel Community. Journal of Travel & Tourism Marketing, 28(7), 675?688. Retrieved from 10.1080/10548408.2011.611739

Levy, S. E., Duan, W., & Boo, S. (2013). An Analysis of One-Star Online Reviews and Responses in the Washington, D.C., Lodging Market. Cornell Hospitality Quarterly, 54(1), 49?63. doi:10.1177/1938965512464513

Linaza, M. T., Agirregoikoa, A., & Garcia, A. (2011). Image-based Travel Recommender System for small tourist destination. Information and communication technologies in tourism.

Luca, M. (2011). Reviews , Reputation , and Revenue?: The Case of Yelp . com. Harvard Business school working paper.

Marchiori, E., Eynard, D., Inversini, A., & Cantoni, L. (2011). Harvesting Online Contents: An Analysis of Hotel Reviews Web sites.

Melián González, S., Bulchand Gidumal, J., & González López-Valcárcel, B. (2010). la participación de los clientes en sitios web de valoración de servicios turístico.El caso de TripAdvisor. Análisis Turístico, (2o semestre), 17?22.

Miguéns, J., Baggio, R., & Costa, C. (2008). Social media and Tourism Destinations?: TripAdvisor Case Study. Advances in Tourism Research 2008, 2008(Porceeding), 1–6.

Minube.com. (2011). Libro blanco de los viajes sociales.Como internet y el protagonismo de los viajeros han revolucionado el sectro turistico. http://www.minube.com/externos/libro_blanco_de_los_viajes_sociales_revolucion_movil.pdf.

Mistilis, N., & Buhalis, D. (2012). Challenges and potential of the Semantic Web for tourism. E-review of Tourism Research, 10(2), 51?55. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=hjh&AN=84339717&lang=es&site=ehost-live&scope=site

Montejo-ráez, A., Perea-ortega, J. M., García-cumbreras, M. Á., & Martínez-santiago, F. (2011). Otium: A web based planner for tourism and leisure Otiu, 38, 10085?10093. doi:10.1016/j.eswa.2011.02.005

O?Mahony, M. P., & Smyth, B. (2009). Learning to recommend helpful hotel reviews. Proceedings of the third ACM conference on Recommender systems - RecSys ?09, 305. doi:10.1145/1639714.1639774

Ong, B. S. (2012). The Perceived Influence of User Reviews in the Hospitality Industry. Journal of Hospitality Marketing & Management, 21(5), 463?485. doi:10.1080/19368623.2012.626743

Ortiz, A. M., Castillo, F. P., & García, R. H. (2010). Análisis de Valoraciones de Usuario de Hoteles con Sentitext *: un sistema de análisis de sentimiento independiente del dominio, 31?39.

Ott, M., Cardie, C., & Hancock, J. (2012). Estimating the Prevalence of Deception in Online Review Communities. Proceedings of the 21st international conference on World Wide Web WWW 12, 201?210. Retrieved from http://dl.acm.org/citation.cfm?doid=2187836.2187864

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding Deceptive Opinion Spam by Any Stretch of the Imagination. Computational Linguistics, 17(June 19-24, 2011), 11. Retrieved from http://arxiv.org/abs/1107.4557

Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining, Proceedings of LREC. 1320?1326.

Park, S.-Y., & Allen, J. P. (2013). Responding to Online Reviews: Problem Solving and Engagement in Hotels. Cornell Hospitality Quarterly, 54(1), 64?73. doi:10.1177/1938965512463118

Peñalver-Martínez, I., L, V. V. S., Segura, C. M. De, Valencia-garcía, R., García-sánchez, F., & Valencia, U. De. (2011). Minería de Opiniones basada en características guiada por Ontologías. Procesamiento del Lenguaje Natural, (46), 91?98.

Popescu, A., Roses, F., & Grefenstette, G. (2009). Mining Social Media to Create Personalized Recommendations for Tourist Visits, (October).

Scott, S. V, & Orlikowski, W. J. (2012). Reconfiguring relations of accountability: Materialization of social media in the travel sector. Accounting, Organizations and Society, 37(1), 26?40. doi:10.1016/j.aos.2011.11.005

Simms, A. (2012). Online User-Generated Content for Travel Planning - Different for different kinds of trips? E-review of Tourism Research, 10(3), 1?10. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=hjh&AN=84339723&lang=es&site=ehost-live&scope=site

Smyth, B., Wu, G., & Greene, D. (2010). Does TripAdvisor Makes Hotels Better?? *, (08), 1–11.

Stankov, U., Laziæ, L., & Dragiæeviæ, V. (2010). The extent of use of basic Facebook user-generated content by the national tourism organizations in Europe, 105–114.

Stringam, B. B., Gerdes, J., & Vanleeuwen, D. M. (2010). Assessing the Importance and Relationships of Ratings on User-Generated Traveler Reviews. Journal of Quality Assurance in Hospitality & Tourism, 11(2), 73?92. doi:10.1080/1528008X.2010.482000

Tham, A., Croy, G., & Mair, J. (2013). Social Media in Destination Choice: Distinctive Electronic Word-of-Mouth Dimensions. Journal of Travel & Tourism Marketing, 30(1-2), 144?155. doi:10.1080/10548408.2013.751272

TripAdvisor. (2013a). Annual Report. SEC, (x), 1?125.

TripAdvisor. (2013b). TripBarometer de TripAdvisor.La mayor encuesta del mundo sobre alojamiento y viajeros. Invierno 2012/2013, 1?38.

Vásquez, C. (2011). Complaints online: The case of TripAdvisor. Journal of Pragmatics, 43(6), 1707?1717. doi:10.1016/j.pragma.2010.11.007

Verma, R., Stock, D., & McCarthy, L. (2012). Customer Preferences for Online, Social Media, and Mobile Innovations in the Hospitality Industry. Cornell Hospitality Quarterly, 53(3), 183?186. doi:10.1177/1938965512445161

Wilson, A., Murphy, H., & Fierro, J. C. (2012). Hospitality and Travel: The Nature and Implications of User-Generated Content. Cornell Hospitality Quarterly, 53(3), 220?228. doi:10.1177/1938965512449317

Wu, G., Greene, D., & Cunningham, P. (2010). Merging multiple criteria to identify suspicious reviews. Proceedings of the fourth ACM conference on Recommender systems - RecSys ?10, 241. doi:10.1145/1864708.1864757

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. Tourism Management, 31(2), 179?188. Retrieved from 10.1016/j.tourman.2009.02.016

Xiang, Z., Gretzel, U., & Fesenmaier, D. R. (2009). Semantic Representation of Tourism on the Internet. Journal of Travel Research, 47(4), 440?453. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=hjh&AN=37615154&lang=es&site=ehost-live&scope=site

Xu, X., Meng, T., & Cheng, X. (2011). Aspect-based extractive summarization of online reviews. Analysis, 968?975. Retrieved from http://portal.acm.org/citation.cfm?id=1982396

Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert Systems with Applications, 36(3), 6527?6535. doi:10.1016/j.eswa.2008.07.035

Yoo, K.-H., & Gretzel, U. (2011). Influence of personality on travel-related consumer-generated media creation. Computers in Human Behavior, 27(2), 609?621. doi:10.1016/j.chb.2010.05.002