



**Universitat de les
Illes Balears**

Title: *Influencing Factors on Hotel Review Ratings*

AUTHOR: *Caroline Thiemermann*

Master's Thesis

Master's degree in Tourism Management and Planning

at the

UNIVERSITAT DE LES ILLES BALEARS

Academic year 2015/16

Date: 02nd September 2016

Author signature

UIB Master's Thesis Supervisor Maria Sard Bauzá

Supervisor signature

UIB Master's Thesis Co-Supervisor (if required)

Co-Supervisor signature

Accepted by the Director of the Master in

Signature

Table of content

Abstract	3
1 Introduction	3
2 Literature Review	5
3 Analysis	8
3.1 Data	8
3.2 Methodology	10
4 Results	12
4.1 Descriptive analysis of the traveler related variables	12
4.2 Analysis of the individual score	12
4.3 Frequency distribution of positive and negative comments according to each evaluation item .	14
4.3.1 Cleanliness.....	14
4.3.2 Hotel complex	15
4.3.3 Location.....	15
4.3.4 Staff	15
4.3.5 Facilities	15
4.3.6 Pool.....	16
4.3.7 Food.....	16
4.3.8 Room	16
4.3.9 Not hotel	17
4.4 Correlation between evaluation items and the individual score	17
4.5 Regression Analysis	19
4.5.1 Individual score as dependent variable.....	19
4.5.2 Sum from valued categories as dependent variable.....	21
4.5.3 Number of valued categories as dependent variable	22
4.6 Discussion	23
5 Conclusion.....	26
6 Bibliography.....	28
Appendix	30

List of tables

Table 1: Description of the variables..... 10
Table 2: Distribution of traveler related variable 12
Table3: Statistical key figures of traveler related variables referred to the individual score 13
Table 4: Two-Sample t-test of traveler related variables assuming unequal variances..... 14
Table 5: Correlations of traveler related variables 18
Table 6: Models 1 – 5 with coefficients and standard error 21
Table 7: Models 6 -7 with coefficients and standard Error 23

List of figures

Figure 1: Evolution of the individual score over the months13

List of annexes

Annex 1: Frequency distribution of positive and negative comments according to traveler related variables 30
Annex 2: Correlation of all traveler related variables32

Abstract

The internet has changed the travel market and online hotel reviews have gained importance both for consumers as well as for suppliers of accommodation services. This study is conducted from the hotel perspective and analyses how online hotel reviews can be used by hotel management to improve their services. It is assumed that the higher the satisfaction level of a guest with different hotel features is, the higher will be the rating of the review and the better will be the hotel performance in terms of sales numbers, which is the final goal. Therefore, the significant features shall be filtered out in order to provide hotels with information about those factors that most influence in the hotel's rating. The study is centered around four-star hotels on the Balearic island Mallorca, a number one beach holiday destination in Europe.

1 Introduction

Travelling has experienced a revolution over the past years as the internet made that tourism demand and supply could be synchronized and managed more easily. The numerous online searching and reservation portals for holiday packages, flights and accommodation open a wide range of selection possibilities to the customers who can autonomously plan and buy their holiday product. With this development, information flows have grown and not only service providers have the possibility to offer information about their product. More and more also users who have already bought the product share their opinion in order to provide potential customers with insights about their experience. Especially online hotel reviews have become very important over the past years due to the appearance of evaluation portals like *tripadvisor*. Today, not only special evaluation websites offer the possibility to exchange opinions, but also the booking portals themselves have integrated the option to publish user-generated content. Having this in mind, the evaluation comments have become essential for service businesses because customers take into account what previous users report about their experience. Hence, reviews are likely to influence the reservation and sales numbers and can potentially determine the success of a business.

This study has the purpose to investigate how hotels can make use of the information provided by the travelers who have stayed in the hotel previously. Due to the fact, hotel reviews can potentially influence the booking behavior of travelers, it is important to manage comments properly. This management can be reactive which would mean that the hotel answers a hotel review that has already been published. However, it could also be proactive by knowing previously which factors influence a comment and its numerical score. The research question

that shall be answered therefore is which factors in an online hotel review have the highest influence on the individual score given by a traveler. This information can help improving operational and through this overall business performance.

As example, the four-star hotel market of Mallorca will be analyzed by help with the hotel reviews from the reservation website *www.booking.com*. Mallorca is the biggest of the Balearic Islands and forms part of the number one tourism destinations for beach holidays in Europe. The touristic sector is the main income source and is currently experiencing an expansion (Pricewaterhouse Coopers, 2014). According to the Institute of Statistics of Balearic Islands, about 8.806.950 tourists arrived in Mallorca in 2015 (Insitut d'Estadística de les Illes Balears, 2016). With a total of 7,747,338 travelers arriving from foreign countries, almost half of them came from Germany accounting for 3,237,745 Germans spending their holidays on the Balearic island. This year, the room nights for July increased by 7.4% compared to the same month in 2015 (Hosteltur, 2016). The growth of the German market accounts for 4.1% from July 2015 to July 2016 (Hosteltur, 2016).

Those numbers show the importance that the German demand has for the island of Mallorca. In 2015, Mallorca counted with 1,639 hotel establishments being around 20% of them four-star hotels or hotel apartments (Govern de les Illes Balears, 2016). The market of four-star hotels is the second biggest one behind the three-star hotels accounting approximately for 30% of the whole accommodation market. The rest is constituted by one-, two- or five-star hotels, agro-tourisms, guest houses, campsites or hostels (ibid., 2016).

The paper is structured as follows: The first part will be a literature review of existing research topics on which this investigation will be based on. At the end of that section, the research question will be defined. Subsequently, the data collection will be explained, followed by the description of the methodology applied for the statistical analysis. This part is followed by the presentation of the results with a discussion of the main findings and some practical implications which might be useful for the industry. The conclusion summarizes the work and will refer to limitations of this study. Moreover, further research topics in relation with this study will be suggested.

2 Literature Review

In the first place the existing literature was reviewed in order to find out what authors had previously found out about the topic. This work is located within the literature on usefulness of online hotel reviews for hotels and their influence on hotel performance. A hotel is a service business which implies that it cannot be tested previously by the consumer. This increases the importance of hotel reviews that guests can publish on booking or review websites in order to share their experiences after their stay with other travelers (Sparks & Browning, 2011). This type of spreading information online is called electronic Word-of-mouth, also known as e-WOM and is characterized by the input of several consumers who already know a product or service and who share their opinion with other interested people (Sparks & Browning, 2011). The possibility of sharing hotel reviews provides travelers with a lot of information about the hotel under consideration. Calveras & Orfila (2014) investigate the role of intermediaries like travel agencies or tour operators in the hotel sector considering the uncertainty about quality with a hotel stay being a product that cannot be completely known before the purchase. They assume that intermediaries can reduce the effect of asymmetric information between the seller and the buyer by transmitting credibly a certain level of quality (ibid.). Although the study is based on travel agencies and tour operators who reduce the information asymmetry for the buyer, this approach could be transferred to online review writers who can also help to increase information flows towards the seller, in this case the hotel.

Several studies deal with the effects that online hotel reviews can have on the hotel performance, purchase intention and product purchase (Sparks & Browning, 2011; Xie et al., 2014; Ye et al., 2009). Sparks & Browning (2011) conduct a study in which they investigate in how far online reviews influence the willingness to book a hotel room and in how far the trust in a hotel can be influenced. One of their findings was that customers rather trust in companies whose employees show much effort to provide a tailor-made customer service. Meaning that employees have a huge potential to influence the trust a customer has into a firm. Another result regarding booking intentions was the fact that recent positive reviews can potentially compensate several negative reviews. This leads to the fact that online reviews can influence the image and popularity of a hotel and consequently its reservation numbers and success (Sparks & Browning, 2011).

Accordingly, Ye et al. (2009) establish a relationship between positive online hotel reviews and the amount of bookings a hotel receives by analyzing China's largest travel website *ctrip*. They

state that according to their results, sales could be increased by 4.4% if the rating score given by reviewers was improved by 10% (ibid., p. 181). Kima et al. (2015) carry out a study with product sales data of an international hotel chain in combination with hotel reviews from “TripAdvisor, Priceline, Hotels.com, Expedia, and Yelp” (Kima et al., 2015, p.167). The authors also confirm the positive relationship that seems to exist between review ratings and the performance of a company.

On a higher level reviews have been analyzed resulting in the conclusion that “review valance, variation and volume positively predict product performance” (Xiea et al., 2014, p.8). Other studies go into more detail and try to find out which factors mostly determine guest satisfaction and review ratings (Xiang et al., 2015; Xiea et al., 2014; Zhoua et al., 2014) Xiea et al. (2014) test in how far the ratings of the features service, room, cleanliness, location and purchase value have an impact on hotel performance, which they measure in product sales. An important result is that the positive effects of the cleanliness and the location rating on a hotel’s performance is stronger for hotels which show high variations in ratings. Opposite to that, the purchase value ratings impacted more negatively in the case of high rating variation. A high number of reviews for one hotel, that is to say the review volume, further enhanced those effects.

Zhoua et al. (2014) study the hotel guest satisfaction in Chinese four- and five-star hotels. For this purpose, they identify the satisfaction features room, hotel, food, value, location and staff with 17 significant subcategories with the aim of finding out which of them has the greatest impact on guest satisfaction. According to their study, features of the hotel room are especially important, as well as the language skills of the staff. Furthermore, facilities as for example the fitness area or the pool are able to notably improve the satisfaction level of the guests. Cleanliness turned out to be an attribute that often caused dissatisfaction.

Xiang et al. (2015) investigate on how hotel guest experience is related with satisfaction. It is done through a big data analysis, collecting 60,648 hotel reviews on the online travel agency website Expedia from more than 10,000 hotels in the US. The goal was to find out the most frequent words that hotel guests use when expressing their satisfaction with a hotel stay. The authors find out that the “head” words center around the core and basic products/services as well as important attributes such as the guest room, cleanliness, staff, location, comfort, service, friendliness and helpfulness of staff, breakfast, bed, and price, etc.” (Xiang et al., 2015, p. 126). Another result of the study is that several valued factors are connected and influence each other.

Thus, in many cases the negative valuation of one also has effects on other factors a hotel guest evaluates. Furthermore Xiang et al. (2015) state that the characteristics of a guest such as the traveling persons can have different effects on the degree of satisfaction.

Research has been done on how subjective data from hotel reviews can be extracted and on the impact of online reviews, especially on how they can influence customers in buying a service. To a quite big extent, researchers have investigated how hotels have to include the online reviews in their customer relationship management. Zhang & Vasquez (2014), for example, investigate the different methods and behaviors that hotels adopt when they answer negative hotel reviews. The study underlines that the way in which hotels answer a hotel review is not only important with regard to the customer who complained but it is also essential with regard to the potential customers who read the reaction of the hotel. Actually, many bigger hotel chains like Hilton, Starwood, and Marriott are using review software which helps them to manage and analyze guest generated content online (Jones, 2010). This approach is rather reactive considering that in this case the hotel has already provided the service.

However, research is still lacking deeper investigation on which drivers determine if a hotel review and its numerical individual score turns out better or worse. Kima et al. (2015) point out the importance of online hotel reviews for hotel management as they can provide managers with ideas for improvement: “By improving the relatively deficient areas in hotel products and services, hotels can expect much better online reviews from customers” (Kima et al., 2015, p. 170).

The numerical score hotels receive can directly and indirectly influence the booking behavior: on one side, consumers tend to book hotels that have better scores because former customers seemed to be more satisfied with the provided services (Canzoniere, 2014). On the other side, the algorithms on hotel booking websites take into account the comment ratings for the positioning of a hotel (Canzoniere, 2014). Thus, the higher the score of a hotel is, the more positive is the impact on the hotel’s ranking. Following this idea, the objective of a hotel should be achieving high global scores which are calculated by the average of the individual scores related to each hotel review. Consequently, it is interesting for a hotel to know, which the main factors that positively or negatively influence the individual score, are. This information extracted from hotel reviews can be used by the hotel in order to improve the operational

performance, the individual review scores, the ranking position, booking intentions and lastly the sales numbers.

One important aspect to have in mind are the difficulties at extracting useful information from subjective evaluations as it is a hotel review (Maharani et al., 2015; Pekar & Ou, 2008). Pekar and Ou (2008) name three levels on which hotel reviews can be analyzed: “the document level, the sentence level and the feature level” (Pekar & Ou, 2008, p. 146). The document level refers to the highest level taking into account the overall opinion of a document. Opinion mining on a sentence level tries to classify whole sentences into positive or negative. And the information extraction on a feature level has the aim to identify how a customer evaluates different features of a product or service. Pekar and Ou (2008) analyze hotel reviews on the feature level by trying to identify and carry out a polarity classification with subjective expressions about location, food, room, services, facilities and price. The objective is to find ways to extract these opinions automatically in order to summarize non-numerical evaluations more easily.

Based on the previous literature review, it can be reasoned that information extraction can improve hotel management, traveler ratings, the positioning on booking websites, the willingness to book and with that the hotel performance in terms of sales numbers. Although some studies have been carried out for similar topics, as far as it could be found out, no study centering on German travelers and their behavior exists.

3 Analysis

3.1 Data

For the purpose of this study, a database of hotel reviews given by travelers on the website *www.booking.com* was especially developed to carry out the analysis. Hotel reviews were taken from 98 four-star hotels located all over Mallorca, covering big hotel complexes on the coast as well as smaller Finca hotels in the interior part of the island. As this analysis was done for the German travelers on the island, only reviews in German language were considered. The website allowed filtering the different languages which made it easy to display only reviews in a certain language. Most of the reviews came from German travelers, however, all the reviews coming from Swiss, Austrian or other travelers were included too, as long as they were written in German. As for the period of time for the contemplation of reviews, one year was determined in order to take into consideration the seasonality effects that might influence hotel reviews over the months. More concise this means that all the reviews between the 11th of June 2015

and the 11th of June 2016 were included in the database. The original list of hotels included 100 hotels, however, one of them was not represented on *www.booking.com* and another one didn't have any German reviews in the chosen time period. In total, 3,166 reviews were listed.

For each hotel, the global score was copied which is the sum of the reviews in all languages and from a broader time period. Then, from each review the following information was collected: the title, the numerical individual score, the written comments under the positive and the negative section as well as traveler related characteristics as the type of trip, the traveling person, the room type, the length of stay, the type of device, the gender as well as the date of the review.

In order to carry out the analysis, not all the hotels were taken into account, but only a smaller sample group. A representative selection was done in the following way: in a first step, the average numerical score for each hotel was calculated based on the collected reviews in order to establish a ranking of hotels according to their individual review scores. This means, all individual scores of one hotel were summed up and divided by the number of reviews. Once established this ranking, the three hotels with the highest score, the three hotels with the lowest score and three hotels with an average score were chosen. The average score has been calculated by summing up all global scores of the hotels and dividing through the number of hotels. Those three hotels were selected which in the ranking had the same average score as the total average of all global scores. Like this, it was assured that hotels with good, average and bad scores are included in the analysis. Once having the nine hotels, the evaluation items appearing in the comments had to be determined. As a basis, the predetermined categories from *www.booking.com* were included: cleanliness, comfort, location, facilities, staff, value for money and WIFI. After having checked the written comments, the list was expanded by the following categories in order to capture all the factors appearing in the reviews: hotel complex, parking, pool, food, price, room, not hotel.

With this list of factors the hotel reviews could be quantified by coding the appearing factors with -1 if the comment about a feature was negative, with 0 if nothing was said about the factor and with +1 if the comment about the factor was positive.

Before starting with the analysis, data was cleaned and those reviews were omitted that didn't have information about the gender or that represented very special room type categories that

didn't have many observations. After this correction, the final sample for the data analysis included 170 hotel reviews from German speaking travelers.

3.2 Methodology

The aim of the analysis was to find out if and in how far the dependent variable individual score of a hotel review is influenced by different independent variables. Due to the great amount, a preselection of all collected independent variables was done and the final result of variables used in the analysis are represented in table 1. It also shows the possible options that later served as dummy variables. From the traveler related characteristics the variable *type of trip* was omitted because the number of observations for *business* was minimal compared to *leisure* and therefore wouldn't have led to meaningful results. With regard to the evaluation items, those features were selected for the analysis that in total counted with an evaluation (+1 or -1) in more than ten reviews. It was assumed that the most mentioned items also were the ones with the highest impact on the comment from a traveler's point of view.

Individual score	Rating given for one review
Traveling person	<i>Single, couple, family or group</i>
Room type	<i>Standard room</i> summarizing standard double or standard triple rooms, <i>superior room</i> summarizing the denominations superior/deluxe/premium room, <i>room with balcony</i> or <i>apartment</i>
Length of stay	<i>Under a week</i> including stays from one to four nights, <i>week</i> including stays from five to 8 nights, and <i>over a week</i> including stays from nine to fourteen nights
Type of device	<i>Mobile "yes"</i> or <i>Mobile "no"</i>
Gender	<i>Female</i> or <i>Male</i>
Cleanliness	Cleanliness in general or in the rooms
Hotel complex	Everything concerning the hotel as a whole as it could be for example the atmosphere, the beauty etc.
Location	Closeness to places like the city center, the beach, the mountains as well as to the location for the purpose of the traveler
Staff	Service, helpfulness and friendliness of the staff members
Facilities	Animation, gym, spa available in the hotel
Pool	Cleanliness, the variety, the sunbeds, of the pool area
Food	Quality, variety and preparation
Room	Everything related with the room, furniture, bathroom, air condition, balcony, view, noise (if specifically mentioned related with the room)
Not hotel	Everything that is not within the influence of the hotel itself as it might be the weather, the village, or noise from people in the street
Sum of valued categories	Sum of all 1,0,-1 within one review
Number of valued categories	Number of categories within one review that have been valued with 1 or -1

Table 1: Description of the variables

For the purpose of the data analysis statistical key figures, t test, correlations and regressions were carried out with the Excel Add-In tool “Data Analysis”.

The statistical key figures were created with the help of the tool “Descriptive Analysis” and will be presented in the following chapter. Furthermore a two sample t test completed the analysis with the aim to see how the individual score variable behaved when contrasting between the subcategories of the traveler related variables. Unequal variances were estimated as these values were not known.

Furthermore, a frequency distribution table was established, which due to its extension can be found in the annex (Annex 1). Annex 1 shows in the first two columns the numbers of observations and frequency that resulted for each traveler related independent variable and its possible subcategories. The following columns contain the frequencies with which the different subcategories mentioned positive (1) or negative (-1) points about an evaluation item or didn't say anything (0).

The next step was the calculation of the correlation coefficients through the option “Correlation” in Excel's Data Analysis tool in order to see the relation between the explaining variables and the individual score. The dummy variable results were included in order to make the analysis more complete.

Subsequently, a regression analysis was carried out in order to find out in how far a cause and effect relationship could be established between the different independent variables and the variable that shall be explained. For those variables that had various specification categories, dummy variables were used. This was the case for the traveler related variables as well as for the evaluation categories (-1, 0, +1). As reference categories for the traveler related variables were chosen the dummies *couple*, *standard room*, *week*, *mobile no* and *male*. For the evaluation categories, the “no comment” category (0) was selected.

Due to the fact the analysis was done with the Excel Add In, a maximum of 16 explaining variables could be considered for one model given that this is the restriction of the program. Considering this, several regression models have been tested to explain the *individual score*, but also the *sum of valued categories* and the *number of valued categories*. Having in mind the restriction of 16 explaining variables as maximum within one model, several options were tried

for each regression trying to substitute the variables with the lowest p-value by other variables in order to achieve the model with the highest possible adjusted R².

4 Results

4.1 Descriptive analysis of the traveler related variables

The characteristics among the travelers in this study were distributed as shown in table 2. In this study more men (58.2%) than women (41.8%) left comments on *www.booking.com*. Most of the travelers were couples (66.5%), followed by families (15.3%), groups (11.8%) and the smallest amount were single travelers (6.5%). As for room types, the distribution among the sample group was relatively equilibrated accounting for 36.5% for a standard room type, 25.9% for a superior room, and almost the same percentage for a room with balcony (19.4%) and an apartment (18.5%). The frequency for length of stay didn't vary a lot between stays of under a week (42.4%) and of a week (41.8%). However, significantly less travelers stayed for more than a week with only 15.9%. In the sample group, less people sent the hotel review through a mobile device accounting for 44.7% compared to a higher amount of travelers who didn't use a mobile device (55.3%).

Variable	Frequency	Variable	Frequency
Traveling Persons	100.0%	Length of Stay	100.0%
Couple	66.5%	Under a week	42.4%
Family	15.3%	Week	41.8%
Group	11.8%	Over a week	15.9%
Single	6.5%	Mobile	100.0%
Room Type	100.0%	Yes	44.7%
Apartment	18.2%	No	55.3%
Balcony	19.4%	Gender	100.0%
Standard	36.5%	Female	41.8%
Superior	25.9%	Male	58.2%

Table 2: Distribution of traveler related variable

4.2 Analysis of the individual score

Figure 1 represents the evolution of the mean of the individual scores for each month. During most of the months the individual score is relatively constant with slight variations between 8.5 and 9 points. Two sharp falls can be observed in December 2015 and February 2016 where the average score drops to approximately 7 points respectively. This might be due to the fact that hotels in winter are not as prepared as in summer and all the services are provided on a reduced

level. The peak in January 2016 between the two drops could be explained by the fact that around Christmas and New Year's Eve the service level is better as also occupation increases in that time. However, in order to recognize a real trend over the months, it would be necessary to observe data for several years.

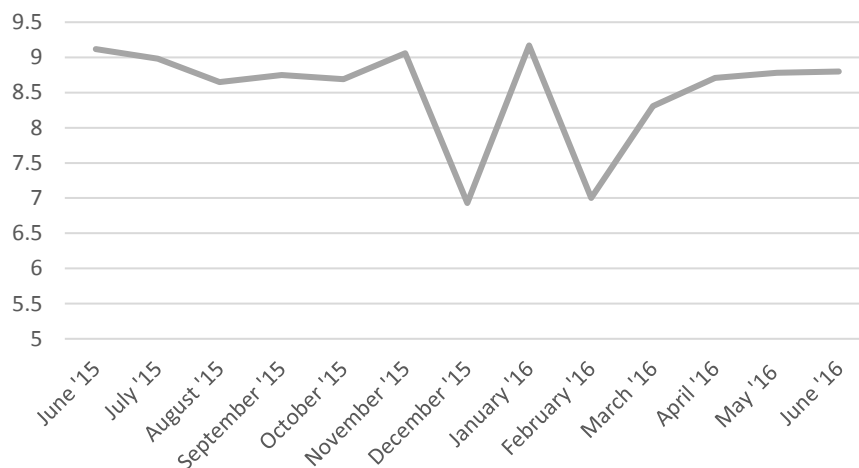


Figure 1: Evolution of the individual score over the months

Table 3 shows the statistical figures for the individual score among the different categories of the traveler related variable. The first row contains the results for the total of individual scores. The following rows show the numbers for each traveler related variable. Without exception, the variables are negatively skewed meaning that values higher than the mean are more common than those below the mean. Travelers tend to give rather higher ratings than lower ratings.

Variable	Mean	Std. Dev.	Smpl. Var.	Skewness
Individual Score	8.676	1.459	2.130	-1.669
Couple	8.603	1.575	2.481	-1.598
Family	8.969	0.859	0.738	-0.367
Group	8.600	1.590	2.528	-1.726
Single	8.882	1.101	1.212	-0.671
Apartment	8.703	1.181	1.394	-1.594
Balcony	8.879	1.201	1.442	-1.970
Standard	8.642	1.521	2.312	-1.456
Superior	8.555	1.731	2.995	-1.683
Under week	8.651	1.275	1.626	-1.427
Week	8.644	1.644	2.702	-1.662
Over week	8.830	1.449	2.099	-2.158
Mobile "yes"	8.578	1.494	2.231	-1.356
Mobile "no"	8.756	1.434	2.057	-1.985
Female	8.639	1.547	2.394	-1.756
Male	8.703	1.400	1.961	-1.596

Table3: Statistical key figures of traveler related variables referred to the individual score

Additionally to that, a two-tailed t-test was carried out in order to see if the means between the different subgroups are different. Table 4 shows the results after having contrasted each possibility of the traveler related variables against the other subcategories. It can be seen that the t stat value is below the t critical two-tail value for all the tested variables. This means that the observed differences between the means are not sufficiently big enough as to say that the differences in the individual score can be attributed to the characteristics of the travelers. It might be the case that for each group not enough observations are available as to observe meaningful differences.

Contrasted means	t Stat	t Critical two-tail	Contrasted means	t Stat	t Critical two-tail
Couple/Family	-1.634	1.995	Standard/Apartment	-0.214	1.992
Couple/Group	0.007	2.056	Standard/Balcony	-0.832	1.990
Couple/Single	-0.768	2.145	Standard/Superior	0.269	1.988
Family/Group	0.938	2.052	Apartment/Balcony	-0.590	1.999
Family/Single	0.235	2.131	Apartment/Superior	0.442	1.993
Group/Single	-0.579	2.052	Balcony/Superior	0.970	1.992
Week/Over week	-0.547	2.006	Mobile "yes"/Mobile "No"	-0.790	1.975
Week/Under week	-0.031	1.978	Female/Male	-0.275	1.977
Over week/Under week	0.563	2.018			

Table 4: Two-Sample t-test of traveler related variables assuming unequal variances

4.3 Frequency distribution of positive and negative comments according to each evaluation item

The most interesting results from the frequency distribution of +1 and -1 evaluations according to each traveler related variable and each evaluation item are described in the following subsections (see Annex 1):

4.3.1 Cleanliness

Cleanliness seemed to be more important to men than for women as the respective percentages both for +1 and -1 were higher among men than among women who in general mentioned less times a point about cleanliness (11.1% and 10.1% for men vs. 5.6% and 7.0% for women). The traveling persons that relatively evaluated cleanliness as most positive were the single travelers (18.2% positive comments) and as most negative the families (15.4% negative comments). Most positive comments came from travelers in a room with balcony (21.2%), as for the negative comments no room type stood out. Regarding the length of stay, travelers who stayed less than a week showed a higher percentage for positive points (12.5%), whereas travelers who stayed more than a week showed a higher percentage for negative points (14.8%).

4.3.2 Hotel complex

As for the *hotel complex*, interesting was that families gave the majority of positive comments (34.6%) and of negative comments (11.5%) compared to the other traveling persons categories. In general, relatively more positive points were mentioned for all lengths of stay (over 23%) than negative points (under 10%). Both women and travelers who sent by mobile mentioned more negative points than men and travelers sending not through a mobile device respectively. The rest of the results didn't seem to show significant differences between the frequencies.

4.3.3 Location

Location was valued much more positive than negative across all traveler related variables as the frequencies were much higher for +1 evaluations (from 28.2% upwards) then for -1 evaluations (from 5.0% downwards). Groups and travelers staying in superior rooms seemed to pay more attention to the location as both the most positive and negative comments about this evaluation category came from them. Men made more positive comments about the location (41.4%) than women (28.2%).

4.3.4 Staff

Staff also received much more positive (between 29.2% and 66.7%) than negative (between 0% and 10%) comments across all traveler characteristics. This result also shows that this evaluation category often received an evaluation (+1,-1) compared to a smaller amount of people who didn't mention staff at all (0). The highest positive percentage was given by travelers who stayed over a week (66.7%) and they also wrote no negative reviews about staff. Families were the travelers with the most positive points about staff (57.7%) compared to the groups who revealed the most negative points (10%). Female travelers seemed to give staff more importance as both the frequencies for positive and negative points about this evaluation item were higher among women than among men (50.7% and 4.2% for women vs. 37.4% and 3.0% for men). That means that men (59.6% not mentioning staff) more often showed an indifferent opinion than women (45.1%).

4.3.5 Facilities

Facilities showed high percentages of "0" meaning that this item wasn't mentioned often. As for the traveling persons, families represented the group both with most positive and negative comments (7.7% respectively) compared with the other traveling persons. The most positive percentage among the room types belonged to the room with balcony (9.1%) and the most negative one to apartment (9.7%), however these percentages didn't differ a lot from the percentages belonging to other room types. Groups and singles didn't mention this item at all.

Travelers staying under a week were most indifferent with respect to facilities as the percentage of “0” was 97.2%. Most positively evaluated the travelers staying for a week (7.0%) and most negatively those staying for over a week (7.4%) compared to the other subgroups of the variable. No significant results could be observed for the variables mobile and gender.

4.3.6 Pool

The evaluation category *pool* received most positive evaluations from the couples accounting for 16.8% compared to the other possibilities of traveling persons. The only group that said something negative about the pool were the families with 3.8% although this is a very small percentage. With regard to the room type, the highest number of positive comments was attributed to standard and superior rooms (16.1% and 15.9% respectively). The only category with negative reviews for pool were standard rooms with 1.6%. With regard to the gender, women mentioned positive points about the pool approximately twice as often as men (18.3% vs. 9.1%). No significant differences could be observed for the variables length of stay and mobile.

4.3.7 Food

The evaluation category *food* received most positive comments by couples (59.3%), however the other travelers followed closely (50.0% for families, 45.5% for singles and 40.0% for groups). Most of the negative points were mentioned by the singles (9.1%), followed by the couples (8.0%), the families (7.7%) and the groups (5.0%). In total, couples demonstrated the lowest percentage of “0”. The room type with the most positive frequency was the standard room (62.9%) and the one with the highest frequency for a negative evaluation the superior room (11.4%). As for the variable length of stay only a significant result could be observed on the -1 observations. Here, the travelers staying for a week mentioned more negative points about food than people who stay under or over a week. With regard to the variable mobile, travelers who sent the review through a mobile device evaluated food slightly more positively (57.9%) than those who didn’t (52.1%). Consistently with this result, more negative points were mentioned by those who didn’t send through a mobile device (10.6%) than those who did (3.9%).

4.3.8 Room

The evaluation category *room* received relatively many positive or negative comments as the percentages of “0” are rather low. Interesting is that room was the evaluation category that overall received the highest number of negative mentions. With regard to the variable traveling

persons, the singles showed the highest percentage of positive comments on room with 45.5%, however the other groups were also close to that number (40.0% among groups; 38.5% among families and 28.3% among couples). Most negative comments about the room came from couples with 19.5%. Another interesting fact is that among the room types, the room with balcony received a significantly higher number of positive comments than the other three room type categories accounting for 45.5% compared to 31.8% for superior rooms, 29.0% for apartments and 27.4% for standard rooms. The highest percentage of negative comments on the room was achieved by the standard room type which is consistent with the low frequency of positive points. With respect to the length of stay, the highest mention quote both for positive and negative comments came from travelers staying more than a week. However, all percentages from the other subcategories were quite closely together. For the variables mobile and gender no significant differences were observed.

4.3.9 Not hotel

The last evaluation category *Not hotel* had an overall low answer quote because high frequencies for “0” mean that no comments were given for this item. The only significant result could be observed for travelers who stayed longer than a week who compared to the other two categories showed the highest number of negative comments regarding this evaluation category (14.8% vs. 2.8% for travelers staying a week and 1.4% for travelers staying under a week).

4.4 Correlation between evaluation items and the individual score

Looking at the correlation coefficients in table 5 for the evaluation items and the individual score, most of the results are positive as expected. This means that in general positive evaluations for an evaluation category also was related to a positive evolution for the individual score.

	Indiv. Score	Cleanliness	H.complex	Location	Staff	Facilities	Pool	Food	Room	Not hotel	Sum categ.	Nr. categ.
Individual score	1.0000											
Cleanliness	0.2290	1.0000										
Hotel complex	0.2899	0.1519	1.0000									
Location	0.0694	0.0840	0.0190	1.0000								
Staff	0.2317	0.1007	-0.0500	0.0792	1.0000							
Facilities	0.1181	-0.1013	0.1612	-0.0703	0.0151	1.0000						
Pool	0.0279	0.0000	0.1243	0.0921	-0.0084	0.0690	1.0000					
Food	0.2944	-0.0662	-0.0010	0.0415	0.1413	0.0158	0.1103	1.0000				
Room	0.3384	0.2094	0.1199	0.0164	0.1197	0.1009	-0.0656	0.0465	1.0000			
Not hotel	0.1484	0.1963	0.1058	0.0595	-0.0808	-0.0035	-0.0205	0.0357	0.0831	1.0000		
Sum valued categories	0.5135	0.4260	0.4452	0.3641	0.4220	0.2087	0.2788	0.4292	0.5539	0.2338	1.0000	
Number valued categories	-0.0049	0.1026	0.1073	0.4124	0.3640	-0.1010	0.2682	0.1752	0.2036	-0.1057	0.4986	1.0000

Table 5: Correlations of traveler related variables

Of the tested variables, the highest correlation shows the variable *sum* with +0.5135 showing that an increase in the sum of evaluations within one hotel review, is related to an increase in the individual score. This result is consistent with the fact that the more the individual evaluation categories are valued positively, the better is the overall satisfaction with the hotel resulting in a higher individual score for the hotel as a whole. The evaluation item with the highest positive correlation is the variable *room* with +0.3384, meaning that this evaluation category has the strongest relation with the individual score. This might mean that a positive comment on the room is very often related with a high individual score. The item *room* is closely followed by the items *food* with +0.2944 and *hotel complex* with +0.2899. Those categories also are positively related with an increase in the individual score. Still important are the evaluation categories *staff* (+0.2317) and *cleanliness* (+0.2290), as well as *not hotel* (+0.1484) and *facilities* (+0.1181). Yet, the latter are already less strongly correlated with the variable individual score meaning that they are not associated to such a big extent to an increase in the individual score as the evaluation items *room*, *food* and *hotel complex*. The other correlation coefficients are very close to 0 which is why they are not considered to have a strong relationship with the variable individual score.

An interesting correlation coefficient resulted from the number of valued categories and the individual score. Although it is quite small with -0.0049, yet it is a negative number. This could be a hint for the fact that the fewer categories are mentioned in the review, the higher is the individual. On the opposite this means that if the number of valued categories goes up, the

individual score moves into the other direction. It could mean that travelers rather mention the negative points than the positive ones and that if they are generally satisfied they do not explicitly mention them.

Table 5 also contains the results for each evaluation category with each other evaluation category. Due to the fact the focus of this work are the influencing factors on hotel reviews and on the score a hotel receives, only the correlation coefficients regarding the individual score are interpreted. Annex 2 contains also the correlation coefficients of the variables that have not been further used in the analysis.

4.5 Regression Analysis

The regression analysis should show which of the previously analyzed traveler related variables and evaluation categories has the biggest impact on the individual score a hotel receives. Furthermore, the variables *sum from valued categories* and *number of valued categories* were tried to explain through the independent variables.

4.5.1 Individual score as dependent variable

The first regression including only traveler related variables showed that barely 2% of the variations in the individual score could be explained by the chosen variables ($R^2=0,019$). None of the variables was significant due to the fact that all p-values were higher than 0.05, which was the chosen confidence interval. This means that the characteristics of the travelers didn't cause meaningful differences that would lead to a higher or lower score. The model could be slightly improved by omitting the dummies for traveling persons. Therefore, in the following models the traveler related variables do not contain the traveling persons.

The second regression that only took into account the evaluation categories resulted better than the first one. The best version of this model that after all testing included the negative dummies for *cleanliness*, *hotel complex*, *food*, *room* and the positive dummy for *staff* could explain around 38% ($R^2=0.402$; adjusted $R^2=0.384$) of the variations in the individual score. The other positive or negative dummies were not significant. This means that the highest effects on the individual score in this study had the negative comments on cleanliness, the hotel complex, the food and the room as well as the positive comments on the staff.

The third regression represented an extension of the first regression analysis, including the traveler related variables, except the traveling persons, and the variable *sum of valued*

categories. Including the *sum of valued categories* as additional variable improved a lot the adjusted R^2 compared to the first model which only took into account the traveler related variables. The omission of the traveling persons led to a slight increase of the model's suitability from an adjusted R^2 of 0.228 to 0.238 ($R^2=0.278$ and $R^2=0.274$). Opposite to the low explaining power of the traveler related variables, the sum of the valued categories appeared to explain some variations in the individual score. This result is a hint for the fact that the more categories are valued for example positively, the higher is the sum coming from the valued categories and the better will be the individual score given by the traveler.

The following model also took into account the traveler related variables except traveling persons plus the variable *number of valued categories*. The small R^2 of 0.012 and the negative adjusted R^2 of -0.037% implied that this variable didn't help at all to explain the variations in the individual score. It lets us conclude that it does not seem important how many categories are valued in order to get a higher or lower individual score by the travelers.

The fifth regression model included on one side the traveler related variables, except the traveling persons and on the other side it contained those dummies for evaluation items that in the second regression model had shown a sufficiently small p-value as to be significant: the negative dummies for *cleanliness*, *hotel complex*, *food* and *room* as well as the positive dummy for staff. Furthermore, the *sum of valued categories* was added. Therefore, after the selection of the variables, this model achieved an R^2 of 0.426 and adjusted R^2 of 0.379 when taking into account the number of included variables. Looking at the difference between the R^2 and the adjusted R^2 , we can see that the higher number of explaining variables doesn't lead to a better model. It is probable that this results to a huge part from the traveler related variables as they didn't show good results when testing them in the first regression model.

Although 38% is not a very good result, it fits the data to some part and it lets us make the conclusion that some evaluation items are more important than other ones. This regression shows that the individual score was mostly influenced by negative comments on cleanliness, the hotel complex, the food and the room. Only the positive comments for staff resulted as important for the explanation of the individual score given by a traveler. Results are shown in table 6. Although previously all variables were significant at 0.05, in this model *over a week*, *staff* and sum of valued categories was not significant.

Model 1		Model 2		Model 3		Model 4		Model 5	
Variables	B	Variables	B	Variables	B	Variables	B	Variables	B
Constant	8.645 (0.303)	Constant	8.980 (0.131)	Constant	8.079 (0.267)	Constant	8.773 (0.379)	Constant	9.039 (0.289)
Single	0.268 (0.476)	Cleanliness (-)	-1.175* (0.312)	Superior	-0.287* (0.264)	Superior	-0.119* (0.307)	Superior	-0.368* (0.243)
Family	0.363 (0.346)	H. complex (-)	-1.12* (0.337)	Balcony	-0.0976* (0.285)	Balcony	0.223 (0.330)	Balcony	-0.001* (0.266)
Group	0.018 (0.381)	Staff (+)	0.520 (0.179)	Apartment	0.103 (0.284)	Apartment	0.022 (0.333)	Apartment	-0.116* (0.268)
Superior	-0.091* (0.308)	Food (-)	-1.744* (0.335)	Under week	-0.064* (0.222)	Under week	-0.0123* (0.260)	Under week	-0.173* (0.205)
Balcony	0.194 (0.334)	Room (-)	-1.293* (0.248)	Over Week	0.077 (0.293)	Over Week	0.208 (0.344)	Over Week	0.095 (0.269)
Apartment	-0.061* (0.353)			Mobile yes	-0.0899* (0.201)	Mobile yes	-0.128* (0.235)	Mobile yes	-0.0675* (0.187)
Under week	0.009 (0.264)			Female	0.001 (0.200)	Female	-0.06* (0.233)	Female	-0.087* (0.185)
Over Week	0.205 (0.344)			Sum valued cat.	0.420 (0.055)	Nr. valued cat.	-0.0207* (0.079)	Cleanliness (-)	-1.004* (0.350)
Mobile yes	-0.128* (0.235)							H. complex (-)	-0.971* (0.376)
Female	-0.060* (0.234)							Staff (+)	0.345 (0.219)
								Food (-)	-1.624* (0.363)
								Room (-)	-1.147* (0.291)
								Sum valued cat.	0.112 (0.075)

*Significant at 0.05

Table 6: Models 1 – 5 with coefficients and standard error

The regression equation with the significant variables would result as follows:

$$\text{Individual score} = 9.039 - 0.368*\text{Superior Room} - 0.001*\text{Balcony Room} - 0.116*\text{Apartment} - 0.173*\text{Under week} - 0.068*\text{Mobile "yes"} - 0.087*\text{Female} - 1.004*\text{Cleanliness (negative)} - 0.971*\text{Hotel Complex (negative)} - 1.624*\text{Food (negative)} - 1.146*\text{Room (negative)}$$

4.5.2 Sum from valued categories as dependent variable

Another regression model has been established in order to find out which variables mostly explain the *sum of valued categories*, which is taken as dependent variable. The main objective was to find out if the characteristics of the travelers might cause differences in that sum. This

model took into consideration the *room type*, *the length of stay*, *the device*, *the gender*, *cleanliness* (positive and negative), *hotel complex* (positive and negative), *location* (positive and negative) and *staff* (positive and negative). The variables fit quite well the data explaining 61% of the data ($R^2=0.606$ and adjusted $R^2=0.567$). However, the traveler related variables (including all except traveling persons) were not significant at 0.05 confidence level and therefore didn't contribute to the explanation of the variations in the dependent variable. For results see table 7 Model 6.

4.5.3 Number of valued categories as dependent variable

The results for the regression of the variable *number of valued categories* against the explaining variables show that the characteristics of the travelers alone are not able to explain the variations in the dependent variable. The more evaluation items are considered in the regression model, the better the results are for the adjusted R^2 . The final model that showed the best result in terms of R^2 and adjusted R^2 (0.791 and 0.770) and smallest p-values contained the following explaining variables: *room type*, *the length of stay*, *the device*, *the gender*, *cleanliness* (positive and negative), *hotel complex* (positive), *location* (positive), *staff* (positive), *facilities* (negative), *pool* (positive), *food* (positive) and *room* (positive). Those evaluation items had received most evaluations of +1 or -1. For results see table 7 Model 7.

Model 6		Model 7	
Variables	<i>B</i>	Variables	<i>B</i>
Constant	0.453 (0.270)	Constant	0.788 (0.180)
Superior	0.304 (0.251)	Superior	-0.143* (0.152)
Balcony	0.555 (0.279)	Balcony	0.029 (0.169)
Apartment	0.260 (0.272)	Apartment	-0.323* (0.165)
Under week	0.100 (0.214)	Under week	-0.061* (0.131)
Over week	-0.217* (0.282)	Over week	0.128 (0.168)
Mobile yes	0.059 (0.193)	Mobile yes	0.118 (0.115)
Female	-0.04* (0.195)	Female	-0.083* (0.119)
Cleanliness (+)	1.029 (0.346)	Cleanliness (+)	1.333 (0.215)
Cleanliness (-)	-1.435* (0.342)	Cleanliness (-)	1.541 (0.206)
H. complex (+)	1.152 (0.217)	H. complex (+)	0.824 (0.130)
H. complex (-)	-1.901* (0.362)	Location (+)	1.143 (0.121)
Location (+)	1.050 (0.203)	Staff (+)	0.861 (0.126)
Location (-)	-1.383* (0.893)	Facilities (-)	2.174 (0.292)
Staff (+)	1.334 (0.210)	Pool (+)	1.216 (0.173)
Staff (-)	-0.901* (0.528)	Food (+)	0.691 (0.118)
		Room (+)	0.878 (0.125)

*Significant at 0.05

Table 7: Models 6 -7 with coefficients and standard error

4.6 Discussion

The previous analysis gave some insights into the behavior of German speaking travelers when writing online hotel reviews. In this part of the study only the most important results are discussed and only those that might be useful for the hotel. The observed differences between the mean individual score for the traveler related variables within this study didn't show significant results in the t test. This means that although certain differences for different characteristics of the traveler could be observed, it was not statistically important as to say that

it is a general behavior of German speaking travelers. A reason for this might be the small amount of observations for the several subcategories in the traveler related variables. This result was also confirmed by the regression analysis where the individual score could hardly be explained by the traveler related variables. Very poor models resulted when only taking into account the characteristics of the travelers.

With regard to the frequencies of the evaluation items some interesting results could be obtained. Cleanliness seemed to be more important for men than for women which can be a useful hint for hoteliers. As more men posted comments on the website *www.booking.com* the point cleanliness could be crucial for the achievement of higher individual scores. According to the results of this study families were the travelers who most complained about cleanliness as well as those travelers who stayed longer than a week. This could be a hint for the fact that the longer a person stays in a hotel, the more negative points are taken into account and are commented with regard to cleanliness.

The results for the valuation item hotel complex showed that families appreciate nice hotel complexes most which could mean that the more a hotel attracts families, the more a hotel should pay attention to this point. Whereas the hotel complex seemed to be especially important for families, the location demonstrated to be very important for groups. Therefore, the hotel should pay special importance to the description of the location on the booking website. It can be supposed that groups come for special purposes like cycling, partying or others. The hotel has the possibility to prevent disappointment and negative comments about the location if the surroundings are well described.

Staff represented one of the variables with the highest answer quote meaning that staff is a point travelers like to mention in their hotel reviews. The fact that personal relationships count a lot in the service business is underlined by the fact that travelers staying more than a week give staff most positive valuations. The longer a guest stays, the more the customer treatment can be personalized and adapted to the needs of the travelers. Then it is more likely that the guest will leave a positive comment. According to the previous analysis, women mentioned staff more often than men in their reviews which might be a hint for a hotel to pay more attention to a personalized treatment for women than for men. However, it should be taken into account that personal treatment is important for all guests. The results also showed that families were the happiest travelers regarding staff. It might be that families require more help from the staff

when children are involved. This is a huge potential for hotels to collect positive points when they know how to satisfy not only the parents but also the children.

The facilities demonstrated to be most important for families. This could be a hint that families make more use of facilities as they stay more time within the hotel due to the children and therefore also consider it important to mention in their reviews. People who stayed less than a week paid less attention to facilities than those staying for one or more week. This might be caused by the fact that those people only stayed for some days and didn't spend much time in the hotel so that they didn't make use of the offered facilities.

With regard to the pool, most comments both positive and negative came from travelers in a standard room. This would mean that a hotel should pay attention to the pool in case it is a hotel with rather many standard rooms. It is interesting that women seemed to mention more positive points about the pool than men, however, it is not clear why this could be the case.

In general, food was very often mentioned positively meaning that this is a factor a hotel definitely should focus in the management. Couples were the traveling persons who most paid attention to food. Another interesting fact was that travelers in the superior room seemed more unsatisfied with the food than those in the standard room. This result might be explained by the fact that travelers who stay in a standard room type have less expectations than travelers who book a superior room type.

The room was the evaluation item that relatively received the most negative comments, which is another result that is useful to know in order to influence review ratings. Interesting was that a balcony was related to a higher frequency of positive comments compared to the other room types. Most negative and least positive comments came from travelers in a standard room which might show that an upgrade to a balcony or superior room can reduce the frequency of negative comments on the room. The slightly higher rate of negative comments from travelers staying for more than week might be due to the fact that the longer a person stays in the room, the more certain things are missed in a temporary living place.

As the initial purpose was to study how the evaluation items and traveler related categories influenced the individual score, the regression results provide us with more information. They helped to explain which of the positive or negative comments in each category could most be

related to the individual score. According to the best regression model (Model 5) that included both the traveler related variables and the evaluation items, the negative comments for *cleanliness*, *hotel complex*, *food* and *room* were those comments that most explained the variations in the individual score. Therefore, the hotels should try their best to avoid negative comments about cleanliness, the hotel complex, the food and the room. The regression coefficients show that the individual score was most affected by negative comments about food and about the room. Cleanliness and the hotel complex had slightly lower coefficients. Combined with the results from the previous practical implications for each of the items, the hotel can manage German speaking guests according to their characteristics purposefully. This would for example mean, that special attention should be paid to those traveler characteristics that demonstrated high frequencies in negative comments for cleanliness, hotel complex, food and room. On the other side, staff could especially focus on those traveler related variables that seemed to bring with them high positive frequencies in order to increase the frequency of positive comments.

Another interesting result was that neither the sum of the valued categories, nor the number of valued categories depended on the traveler related variables (Models 6 and 7).

5 Conclusion

The aim of this study was to find out which factors have major influence on the individual score of online hotel reviews. The analysis was centered on the hotel reviews by German travelers of four-star hotels in Mallorca. The study contributed to finding out those factors that most influenced German travelers in the decision about the numerical score they gave a visited hotel.

It could be found out that the most influencing factors in this study were negative comments about the food, the room, the cleanliness and the hotel complex. Those features should be in the focus of hotel management because they can provide an opportunity to influence the hotel rating of German speaking travelers. Furthermore, several differences according to the traveler related sub-variables could be found out. Most of them resulted for the traveling persons or for the room type, less for the gender and the length of stay. The fact that travelers sent through a mobile device was not relevant. With regard to the existing literature, the results of this study have contributed new insights about the influencing factors on the review score from German travelers especially in four-star hotels.

The study has several limitations that have to be considered. In a first place, the data that was analyzed was only a small part of the whole collected data. It is probable that with a bigger amount of observations the results especially for the traveler related variables would be different. Furthermore, the results from the analysis could not be extrapolated due to the fact that basically descriptive analysis was used. The study was focused on German hotel reviews, which is why travelers with different nationalities might put a different focus on the analyzed evaluation items. This could also result in different results for the characteristics of the travelers because distributions probably change according to the selected countries. Another fact that is limiting the results of the paper is the focus on four-star hotels, which implies that the results for other hotel categories might be different.

For further research it would be interesting to investigate if in a broader data set, variations in the individual score can be attributed to some characteristics of the travelers. Moreover, it could be useful to test if other nationalities tend to show the same behavior in writing online hotel reviews as Germans in order to see if the information can be used for hotel management in general. Another interesting variation would be doing the same analysis for online hotel reviews from several websites at them same time like Expedia, Tripadvisor and Hotels.com in order to take into account potential differences that might exist between the customers of the several websites.

6 Bibliography

- Calveras, A., & Orfila, F. (2014, August). Intermediaries and quality uncertainty: evidence from the hotel industry. *Tourism Economics*, pp. 727-756.
- Canzoniere, F. (2014, December 18). *Tnooz - How to crack the Booking.com algorithm*. Retrieved August 27, 2016 from <https://www.tnooz.com/article/crack-the-booking-com-algorithm/>
- Govern de les Illes Balears. (2016). *Estadísticas del turismo - Alojamientos turísticos*. Retrieved August 28, 2016, from <http://www.caib.es/sacmicrofront/contenido.do?mkey=M10072911244127834137&lang=ES&cont=33206>
- Hosteltur. (2016, August 23). *Las pernoctaciones hoteleras suben un 7,4% aunque baja la estancia media*. Retrieved August 27, 2016, from http://www.hosteltur.com/117666_pernoctaciones-hoteleras-suben-74-baja-estancia-media.html
- Institut d'Estadística de les Illes Balears. (2016). *Ibestat - Institut d'Estadística de les Illes Balears*. Retrieved August 27, 2016 from http://www.ibestat.es/ibestat/estadistiques/043d7774-cd6c-4363-929a-703aaa0cb9e0/ef88f7cf-8e0b-44e0-b897-85c2f85775ec/es/I208002_3001.px on 30.08.2016
- Ivar E. Vermeulen*, D. S. (2009, February). Tried and tested: The impact of online hotel reviews. *Tourism Management* (30), pp. 123 - 127
- Jones, B. (2010, March 22). *USA Today*. Retrieved August 28, 2016 from Hotel managers monitor online critiques to improve service: http://usatoday30.usatoday.com/tech/news/2010-03-23-businessstravel23_ST_N.htm
- Kima, W. G., Limb, H., & Brymerc, R. A. (2015). The effectiveness of managing social media on hotel performance. *International Journal of Hospitality Management* (44), 165 - 171.
- Maharani, W., Widyanoro, D. H., & Khodra, M. L. (2015). Aspect Extraction in Customer Reviews Using Syntactic Pattern. *Procedia Computer Science* (59), pp. 244 - 253.
- Pekar, V., & Ou, S. (2008). Discovery of subjective evaluations of product features in hotel reviews. *Journal of Vacation Marketing* 14 (2), pp. 145-155.
- Pricewaterhouse Coopers. (2014). *Cámara de comercio de Mallorca - Mallorca, más allá del sol y playa*. Retrieved August 27, 2016, from http://www.cambramallorca.com/documentos/Desp_5293.pdf
- Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, 1310-1323.
- Xiang, Z., Schwartz, Z., Jr., J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest. *International Journal of Hospitality Management*, 120-130.
- Xiea, K. L., Zhangb, Z., & Zhangb, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management* (43), 1 - 12.
- Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management* (28), pp. 180 - 182.

Zhang, Y., & Vásquez, C. (2014, 6). Hotels' responses to online reviews: Managing. *Discourse, Context and Media*, pp. 54-64.

Zhoua, L., Yea, S., Pearceb, P. L., & Wua, M.-Y. (2014). Refreshing hotel satisfaction studies by reconfiguring customerreview data. *International Journal of Hospitality Management* (38), 1 - 10.

Appendix

Variable	Observations	Frequency	Cleanliness			Hotel complex			Location			Staff			Facilities		
			1	0	-1	1	0	-1	1	0	-1	1	0	-1	1	0	-1
Traveling Persons	170	1.000															
Couple	113	0.665	0.071	0.841	0.088	0.283	0.655	0.062	0.363	0.628	0.009	0.398	0.575	0.027	0.035	0.920	0.044
Family	26	0.153	0.077	0.769	0.154	0.346	0.538	0.115	0.308	0.692	0.000	0.577	0.385	0.038	0.077	0.846	0.077
Group	20	0.118	0.150	0.800	0.050	0.050	0.850	0.100	0.400	0.550	0.050	0.400	0.500	0.100	0.000	1.000	0.000
Single	11	0.065	0.182	0.818	0.000	0.273	0.636	0.091	0.364	0.636	0.000	0.455	0.545	0.000	0.000	1.000	0.000
Room Type	170	1.000															
Apartment	31	0.182	0.097	0.774	0.129	0.161	0.710	0.129	0.355	0.645	0.000	0.355	0.581	0.065	0.032	0.871	0.097
Balcony	33	0.194	0.212	0.667	0.121	0.273	0.606	0.121	0.303	0.697	0.000	0.576	0.424	0.000	0.091	0.879	0.030
Standard	62	0.365	0.032	0.919	0.048	0.306	0.645	0.048	0.371	0.629	0.000	0.323	0.645	0.032	0.032	0.919	0.048
Superior	44	0.259	0.068	0.841	0.091	0.273	0.682	0.045	0.386	0.568	0.045	0.523	0.432	0.045	0.000	1.000	0.000
Length of Stay	170	1.000															
Under a week	72	0.424	0.125	0.833	0.042	0.292	0.639	0.069	0.319	0.681	0.000	0.292	0.667	0.042	0.014	0.972	0.014
Week	71	0.418	0.056	0.831	0.113	0.239	0.662	0.099	0.380	0.592	0.028	0.479	0.479	0.042	0.070	0.873	0.056
Over a week	27	0.159	0.074	0.778	0.148	0.259	0.704	0.037	0.407	0.593	0.000	0.667	0.333	0.000	0.000	0.926	0.074
Mobile	170	1.000															
Yes	76	0.447	0.066	0.829	0.105	0.263	0.618	0.118	0.342	0.645	0.013	0.434	0.553	0.013	0.026	0.947	0.026
No	94	0.553	0.106	0.819	0.074	0.266	0.691	0.043	0.372	0.617	0.011	0.426	0.521	0.053	0.043	0.904	0.053
Gender	170	1.000															
Female	71	0.418	0.056	0.873	0.070	0.268	0.620	0.113	0.282	0.704	0.014	0.507	0.451	0.042	0.028	0.930	0.042
Male	99	0.582	0.111	0.788	0.101	0.263	0.687	0.051	0.414	0.576	0.010	0.374	0.596	0.030	0.040	0.919	0.040

Annex 1: Frequency distribution of positive and negative comments according to traveler related variables

Variable	Observations	Frequency	Pool			Food			Room			Not Hotel		
			1	0	-1	1	0	-1	1	0	-1	1	0	-1
Traveling Persons	170	1.000												
Couple	113	0.665	0.168	0.832	0.000	0.593	0.327	0.080	0.283	0.522	0.195	0.009	0.938	0.053
Family	26	0.153	0.077	0.885	0.038	0.500	0.423	0.077	0.385	0.538	0.077	0.000	0.962	0.038
Group	20	0.118	0.000	1.000	0.000	0.400	0.550	0.050	0.400	0.550	0.050	0.000	1.000	0.000
Single	11	0.065	0.091	0.909	0.000	0.455	0.455	0.091	0.455	0.455	0.091	0.000	1.000	0.000
Room Type	170	1.000												
Apartment	31	0.182	0.065	0.935	0.000	0.452	0.516	0.032	0.290	0.645	0.065	0.000	0.935	0.065
Balcony	33	0.194	0.091	0.909	0.000	0.515	0.394	0.091	0.455	0.455	0.091	0.000	0.970	0.030
Standard	62	0.365	0.161	0.823	0.016	0.629	0.306	0.065	0.274	0.452	0.274	0.000	0.952	0.048
Superior	44	0.259	0.159	0.841	0.000	0.523	0.364	0.114	0.318	0.591	0.091	0.023	0.955	0.023
Length of Stay	170	1.000												
Under a week	72	0.424	0.125	0.875	0.000	0.569	0.389	0.042	0.292	0.542	0.167	0.000	0.986	0.014
Week	71	0.418	0.141	0.845	0.014	0.521	0.366	0.113	0.338	0.535	0.127	0.014	0.958	0.028
Over a week	27	0.159	0.111	0.889	0.000	0.556	0.370	0.074	0.370	0.444	0.185	0.000	0.852	0.148
Mobile	170	1.000												
Yes	76	0.447	0.118	0.868	0.013	0.579	0.382	0.039	0.329	0.487	0.184	0.000	0.947	0.053
No	94	0.553	0.138	0.862	0.000	0.521	0.372	0.106	0.319	0.553	0.128	0.011	0.957	0.032
Gender	170	1.000												
Female	71	0.418	0.183	0.817	0.000	0.549	0.394	0.056	0.282	0.549	0.169	0.000	0.958	0.042
Male	99	0.582	0.091	0.899	0.010	0.545	0.364	0.091	0.354	0.505	0.141	0.010	0.949	0.040

Annex 1: Frequency distribution of positive and negative comments according to traveler related variables

	Individual score	Cleanliness	Hotel complex	Location	Staff	Wifi	Comfort	Facilities	Value for money
Individual score	1								
Cleanliness	0,22903919	1							
Hotel complex	0,28991152	0,15194529	1						
Location	0,0694197	0,08399356	0,01901852	1					
Staff	0,2316835	0,1007003	-0,04995306	0,07923321	1				
Wifi	1,6552E-17	0,12909944	0,14712039	0,05421761	-0,04875132	1			
Comfort	0,0699694	0,18311355	-0,02618658	-0,05337889	0,08379167	0	1		
Facilities	0,11809777	-0,10129686	0,1611589	-0,07031846	0,01507592	0,09808026	0,00163666	1	
Value for money	0,04715392	0,06253593	-0,03297767	-0,05901468	-0,07306822	0	-0,01818716	-0,08998991	1
Parking	-0,0528336	0,08208538	0,027146	0,06610746	0,11268534	0,15895766	-0,00795756	0,00220103	-0,0244586
Pool	0,02793581	0	0,12426775	0,0920576	-0,00843923	0	-0,02742671	0,0689975	0,06735047
Food quality	0,294436	-0,06624941	-0,00098689	0,04146116	0,14127535	-0,04276381	-0,05708783	0,01579025	0,03168153
Price	0,09224259	-0,09128709	-0,06935322	0	0	0	0	-0,13870643	0
Room	0,33836591	0,20935405	0,11994377	0,01644658	0,11970245	0,04054124	0,0953879	0,10085756	-0,02102445
Not hotel	0,14843334	0,1962642	0,10583681	0,05947504	-0,08079933	0	0,01268422	-0,0035084	0,03898662
Sum valued categories	0,51349474	0,42602917	0,44523927	0,36411356	0,42196952	0,22500053	0,09661772	0,20866913	0,09351816
Number valued categories	-0,0048946	0,10259967	0,10729269	0,4123598	0,3639894	0,16255916	0,05786918	-0,10104021	0,07287667

Annex 2: Correlation of all traveler related variables

	Parking	Pool	Food	Price	Room	Not hotel	Sum valued categories	Number valued categories
Individual score								
Cleanliness								
Hotel complex								
Location								
Staff								
Wifi								
Comfort								
Facilities								
Value for money								
Parking	1							
Pool	0,0626446	1						
Food quality	-	0,11025361	1					
Price	0,13115438	-	0,12095432	1				
Room	0,22480008	-0,1106859	0,04650389	0,05733397	1			
Not hotel	0,179835	0,06555109	-					
Sum valued categories	0,01705809	0,02053083	0,03569283	0	0,0831442	1		
Number valued categories	0,20623403	0,27877762	0,42924905	0,04242651	0,55386138	0,23376226	1	
	0,26157535	0,2682303	0,17519973	0	0,2035858	-0,1057443	0,49861791	1

Annex 2: Correlation of all traveler related variables