



## Volume 34, Issue 4

### Explaining differences in recommendation rates: the case of South Cyprus hotels

Evgeny A. Antipov

*National Research University Higher School of  
Economics*

Elena B. Pokryshevskaya

*National Research University Higher School of  
Economics*

#### Abstract

In this study we demonstrate how publicly available data can be used to work out the indirect importance of various hotel attributes for their visitors. We apply Shapley value decomposition of the recommendation rate to compute the percentage contributions of various attributes to the overall loyalty, which helps us explain why some of Cyprus hotels have higher satisfaction ratings than the others. It appeared that satisfaction with gastronomy is the key driver of the overall satisfaction with Cyprus hotels. We conduct importance-performance analysis for one of the hotels to demonstrate a strategic management application of our empirical analysis. Directions for other potential studies using user-generated content are proposed.

---

**Citation:** Evgeny A. Antipov and Elena B. Pokryshevskaya, (2014) "Explaining differences in recommendation rates: the case of South Cyprus hotels", *Economics Bulletin*, Vol. 34 No. 4 pp. 2368-2376.

**Contact:** Evgeny A. Antipov - eantipov@hse.ru, Elena B. Pokryshevskaya - epokryshevskaya@hse.ru.

**Submitted:** September 06, 2014. **Published:** November 06, 2014.

## 1. Introduction

Nowadays most hotel booking services provide their users with reviews posted by hotel visitors, which gives a wealth of free information to both academics and industry analysts. Not accidentally in the area of hospitality research there seem to be a rising interest to studies that use publicly available data effectively (Chen & Rothschild, 2010; Ye, Law, & Gu, 2009).

Besides a textual review, users rate hotels on several attributes as well as give an overall rating, which gives an opportunity for hotel management to do research on key drivers of hotel visitors' satisfaction. Even though some hotels survey their visitors, they usually do not survey the visitors of other hotels. Despite the existence of studies related to studying satisfaction with hotels (Choi & Chu, 2001; Martin, 1995; Nadiri & Hussain, 2005; Pender & Sharpley, 2004), they have some limitations that we would like to overcome:

1. Most previous research studies were based on proprietary survey data, whereas we use publicly available data that gives an opportunity to track the market situation more frequently than it is usually possible with expensive primary data.
2. Most hotel satisfaction research are case studies considering only one or several hotels. Using data from hotel booking websites allows not only increasing the number of hotels that are included in the sample, but even conducting cross-country research, which may be interesting both for academics and for hotel chains operating in different countries.
3. Measuring indirect importance of service attributes is still a challenging task in market research. In previous research no attempt was made to measure the percentage contribution of various hotel features to the overall customer satisfaction. Some of the most advanced measures of importance actually used in previous research were standardized regression coefficients (Gustafsson & Johnson, 2004), correlation coefficients (Fontenot, Henke, Carson, & Carson, 2007). Recently we have shown that Shapley value approach to the decomposition of regression's R-squared fit measure is a theoretically sound technique which gives attribute contributions that are stable across random samples of data. It also has a relatively high discriminating power, which helps avoiding the problem of placing equal importance on various attributes which is especially common when stated importance is used instead of indirect importance (Pokryshevskaya & Antipov, 2014).

Even though Cyprus is a major tourist destination in the Mediterranean (Pender & Sharpley, 2004), there have been very few customer satisfaction studies that were based on data from Cyprus. The only one that we are aware of aimed at validating perceived service quality measurement scale (Nadiri & Hussain, 2005), but not at figuring out how important various service attributes are. We use Shapley value decomposition of the recommendation rate to work out the percentage contributions of various attributes to the overall satisfaction, which helps us explain why some of Cyprus hotels have high satisfaction ratings while others – low. Then we conduct importance-performance analysis for one of the hotels to demonstrate a strategic management application of our empirical analysis.

## 2. Data

The data on 121 South Cyprus hotels was collected from Holidaycheck.com in August 2014. This website has one the largest number of hotel attributes we were able find on the Internet. A unique feature of Holidaycheck.com as a data source is that within each category users are asked to evaluate a number of items, so that they clearly understand what exactly they

evaluate. So the survey that underlies hotel ratings is very detailed, even though the end user can see only an aggregated rating for each category.

**Category 1: The hotel in general**

- Condition of the Hotel
- General cleanliness of the different areas
- Family friendliness
- Disabled-friendly

**Category 2: Location and surroundings**

- Shopping facilities in the area
- Transportation connections and available excursions
- Restaurants and bars in the area
- Other leisure activities available
- Distance to the beach

**Category 3: Service**

- Friendliness and helpfulness
- Staff's knowledge of foreign language
- Reception, check-in & check-out
- Competence (handling complaints)

**Category 4: Gastronomy**

- Variety of food and drinks
- Quality of food and drinks
- Atmosphere & furnishings
- Overall cleanliness in the restaurant and in the dining areas

**Category 5: Sport, entertainment and pool**

- Leisure activities (e.g. sauna, tennis, entertainment, etc.)
- Overall condition of the pool area
- Quality of the beach
- Child care or playground

**Category 6: Room**

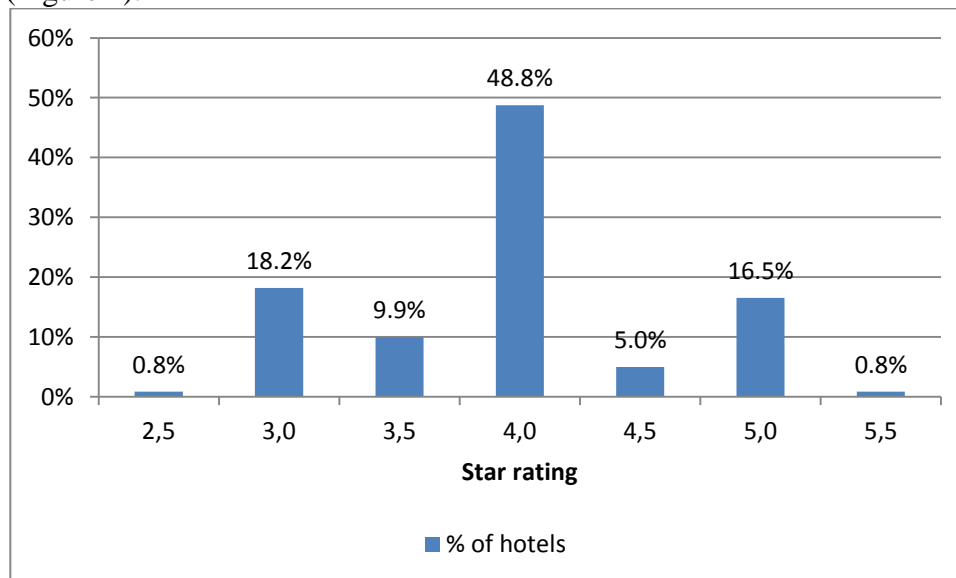
- Cleanliness & linen change
- Size of the room
- Room amenities (TV, balcony, safe, etc.)
- Size of the bathroom

All category ratings are at 6-point scale from 1 to 6. If a feature can't be assessed by a respondent, she or he can choose not to rate it. For example, if a person travelled without kids. At the end of the survey users are asked to answer the question "Would you recommend this hotel?" which gives the percentage of visitors that recommend the hotel. We used this percentage as the *dependent variable*. Category ratings, the number of stars a hotel has and the town in which the hotel is situated were used as *explanatory variables*. Descriptive statistics of scale variables that are used in our analysis are presented in Table 1.

**Table 1. Descriptive analysis of scale variables**

Variable	Definition	Mean	Standard Deviation	Minimum	Median	Maximum
Recommendation rate	% of visitors recommending the hotel	81.5	18.8	0.0	87.0	100.0
Stars	Hotel's star rating	4.0	0.7	2.5	4.0	5.5
General	Satisfaction with the hotel in general	4.9	0.6	2.3	5.0	5.9
Room	Satisfaction with room	4.9	0.6	2.4	5.0	6.0
Service	Satisfaction with service	5.0	0.6	2.6	5.1	6.0
Location	Satisfaction with hotel's location	4.9	0.6	2.5	5.0	6.0
Gastronomy	Satisfaction with gastronomy	4.8	0.7	1.8	4.9	5.9
Sport/Leisure	Satisfaction with sport, entertainment and pool	4.7	0.8	0.0	4.8	5.9

The distribution of hotel star rating is symmetric with almost half of all hotels having a 4-star rating (Figure 1).

**Figure 1. Frequency distribution of hotels by star rating**

### 3. Empirical analysis

Our preliminary analysis has shown that differences across towns are negligible, which is why we do not account for geographic differences in our model. We regressed recommendation rate on 6 attribute ratings and star rating, as well as calculated Cook's distance that is commonly used for detecting outliers (Chatterjee & Hadi, 1986). If Cook's distance exceeds  $4/n$ , where  $n$  is the number of cases in the sample, then the observation is considered to be influential. A less conservative rule of thumb is to pay attention to observations for which Cook's distance exceeds

1. In our sample there are 7 observations with Cook's distance  $>4/n$  and 1 observation with Cook's distance  $>1$ . We have built a robust regression, which puts a smaller weight on influential observations than on normal ones (Rousseeuw & Leroy, 2005). Its parameter estimates are given in Table 2.

**Table 2. Parameter estimates (robust regression)**

	Recommendation rate	
Stars	-8.015***	(1.373)
General	3.943	(3.574)
Room	8.156***	(2.652)
Service	7.437***	(2.618)
Location	-2.943*	(1.720)
Gastronomy	7.771***	(2.481)
Sport/leisure	4.361*	(2.231)
Constant	-26.476***	(8.449)
<i>N</i>	121	
$R^2$	0.798	
adj. $R^2$	0.785	

Standard errors in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

The only insignificant (at 10% significance level) regressor is the rating of the hotel in general. A possible explanation is that people do not pay that much attention to the sub-attributes that comprise hotel condition in general. Every additional point to the average assessment of satisfaction with a room leads to a 8.2 percentage points increase in the probability of recommending the hotel. One point difference in service satisfaction leads to 7.4 percentage points difference in recommendation rate. The effect of satisfaction with food is similar in magnitude, while sport, entertainment and pool have a slightly weaker effect. It is interesting that, other things equal, the higher the star-category of a hotel, the lower its recommendation rate is. This means that people are more likely to recommend a 3-star hotel than a 5-star hotel with the same service. Probably this is explained by the lower value for money in the case of a more expensive hotel that does not fulfill the expectations of its visitors. Negative influence of location rating is surprising, but significant only at 10% significance level and can be a result of omitted variable bias. For example, good location can be correlated with noise, proximity to a road or some unobserved hotel characteristics.

Even though regression results allow doing what-if analysis (e.g. "What if we manage to increase our visitors' satisfaction with food by 0.5 points out of 6?"), we still do not know the relative importance of the attributes. The importance can be defined as the share of explained variance that is attributed to a certain explanatory variable. According to game-theoretic Shapley value approach applied to econometrics, in order to obtain a variable's contribution we have to consider all possible subsets of regressors. In our case this is the number of  $k$ -combinations from a set of 7 regressors, where  $k=1, \dots, 7$ , which equals 127. Shapley value decomposition of  $R^2$  gives us a theoretically sound partitioning of the explained variance. Table 3 reports percentages of Shapley value contributions.

**Table 3. Shapley value contributions**

Explanatory variable	Shapley value contribution, %
Stars	4.86
General	19.28
Room	14.37
Service	20.19
Location	4.15
Gastronomy	21.73
Sport/leisure	15.42

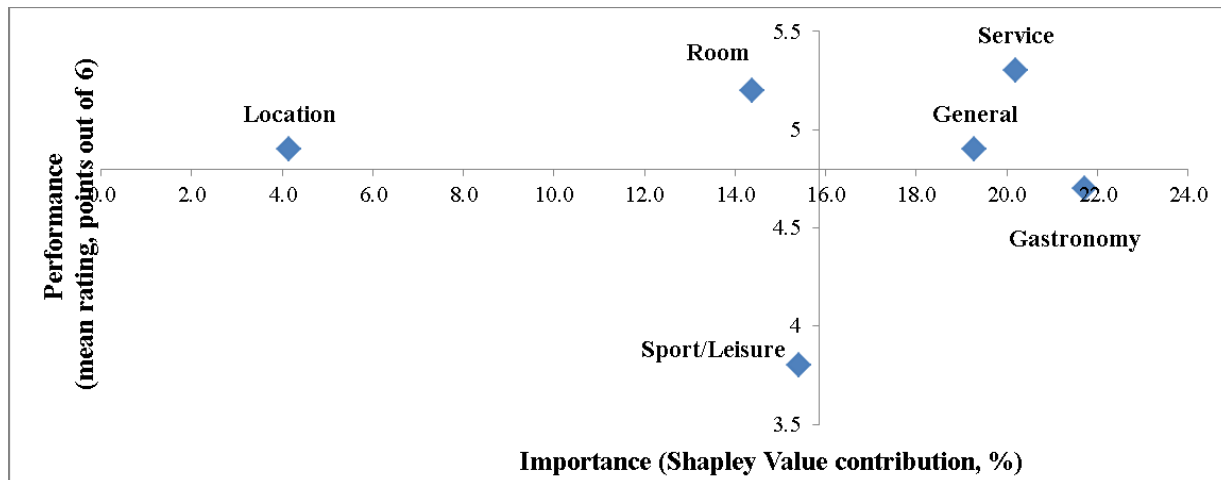
A surprising result is an important role of general hotel ratings (19.28%) despite the fact that this attribute was statistically insignificant according to our regression estimates. A possible explanation for this is high multicollinearity. Even though VIF (Variance Inflation Factor) for our set of regressors does not exceed 10, its value for “General” variable is 7.86, which means that it is highly correlated with some of the other regressors. In particular, this regressor is highly correlated with “Room”, “Service”, “Gastronomy” and “Sport/Leisure” (Table 4). Multicollinearity leads to inflated standard errors and wrong conclusions about statistical significance. We believe the problem of insignificance would disappear if we had a larger sample of hotels, so we tend to believe that the importance of general hotel evaluation is actually higher than one could think based solely on the results of regression analysis. Location appeared to be the least important factor, so we can neglect the fact that its effect is negative (at 10% significance level) according to our regression analysis.

**Table 4. Pearson pairwise correlations between regressors**

	General	Stars	Room	Service	Location	Gastronomy	Sport/Leisure
General	1.000						
Stars	0.428	1.000					
Room	0.839	0.449	1.000				
Service	0.868	0.351	0.763	1.000			
Location	0.478	0.054	0.479	0.495	1.000		
Gastronomy	0.882	0.338	0.816	0.828	0.502	1.000	
Sport/Leisure	0.811	0.395	0.753	0.743	0.567	0.758	1.000

All coefficients are statistically significant at 1% significance level

Attributes that contribute the most to the share of people who recommend the hotel are “Gastronomy” and “Service”. Using Shapley value contributions as importance measures, we did importance-performance analysis for one of the hotels to demonstrate how Shapley value contributions can aid in strategic analysis of a hotel’s strengths and weaknesses. The hotel is situated in Limassol, belongs to a 5-star category and was recommended by 67% of users. We want to figure out the ways to improve the recommendation rate. The strategic quadrant analysis chart is presented in Figure 2.



**Figure 2. Importance-Performance graph**

The description of quadrants is given below:

- Quadrant 1 (high importance, high performance): attributes from this quadrant are the key drivers of satisfaction. It is recommended to maintain the satisfaction with these attributes at a high level. The hotel's visitors are relatively highly satisfied with such important features as service quality and hotel in general. The hotel management should maintain these key drivers of loyalty.
- Quadrant 2 (low importance, high performance): attributes from this quadrant are given too much attention. Rooms and location belong to this segment. It is recommended to avoid spending too many resources on their development.
- Quadrant 3 (low importance, low performance): attributes from this quadrant are minor shortcomings, improving which is not of high priority. It is recommended to delay their improvement up until some point of time in the future. The improvement of sport, entertainment and pool can be delayed.
- Quadrant 4 (high importance, low performance): attributes from this quadrant require immediate improvement. Gastronomy is the only attribute from this category.

By high/low in the definition of quadrants given above we meant higher/lower than averages across all attributed for this particular hotel. Such distinction is simple, but often dissatisfactory: sometimes attributes lie close to the threshold and it is unclear, which segment they belong to. In our case average market performance coincides with the within-hotel average, but if a hotel has relatively low ratings on all attributes, using its own average performance would be a too optimistic approach. This problem can be partly solved by correcting the thresholds using expert judgments and/or market averages. For example, a hotel can have an internal standard according to which high satisfaction with an attribute is attained only if the rating exceeds 5 out of 6 or is higher than 75<sup>th</sup> percentile in the market.

#### 4. Conclusion

Our analysis was intended to explain the differences among South Cyprus hotels at an aggregate level. Shapley value decomposition of robust regression's  $R^2$  has shown that the satisfaction with gastronomy and service are the strongest drivers of customer loyalty measured by whether they would recommend a hotel or not. Other things equal, the higher the star-category of a hotel, the lower its recommendation rate is, which means that a hotel should

consider all pros (the ability to charge higher prices) and cons (lower recommendation rate) of moving from a 4-star category to a 5 star category, from a 3-star category to a 4-star category, etc. If the quality of all attributes remains the same, the hotel is predicted to lose 8 percentage points of recommendation rate, which probably reflects the fact that customers consider value for money when they decide whether to recommend a hotel or not.

In our preliminary research we have not accounted for the price category of the hotel. Price is usually correlated with the number of stars, so we had a relatively good proxy. However, we believe that it is important to account for the pricing in our future research.

One of the directions for future research is explaining, why some people are more satisfied with a hotel than others using not only attribute ratings, but also their personal characteristics. It may be worth accounting for the heterogeneity in the importance of various attributes across different nations. For example, some nations can be more fitness-oriented or food-oriented than others which may influence the emphasis they put on different attributes of hotel service. To do this it is necessary to account for the dependence of model parameters on the country where the reviewer is from.

Hotel booking web services collect data that is invaluable for hoteliers. Taking into account that some of the services ask respondents about their satisfaction with very specific aspects of hotel service, such websites could sell detailed ratings to hotels. This would allow finding more specific drivers of satisfaction within each category. For example, using publicly available data we managed to figure out that the quality of gastronomy is the key driver of loyalty for South Cyprus hotels, but we cannot tell, what should be improved first at a hotel's restaurant: variety, quality, atmosphere and furnishing or cleanliness. HolidayCheck.com collects such detailed ratings in order to compute the publicly available aggregated ratings and could theoretically offer data services to hotels.

Although utilizing user-generated content aggregated by special websites is a promising direction in hospitality and tourism research, online ratings may be biased, i.e. mostly positive and negative reviews are posted, while there are unusually few neutral online ratings (Hu, Pavlou, & Zhang, 2007; Hu, Zhang, & Pavlou, 2009). Another limitation of using freely available data is that predefined (i.e. created by the developers of a web service) measurement scales have to be used. Another limitation is a relatively small number of hotels for which sufficient number of reviews is available.

## 5. References

1. Chatterjee, S., & Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression. *Statistical Science*, 379–393.
2. Chen, C.-F., & Rothschild, R. (2010). An application of hedonic pricing analysis to the case of hotel rooms in Taipei. *Tourism Economics*, 16(3), 10.
3. Choi, T. Y., & Chu, R. (2001). Determinants of hotel guests' satisfaction and repeat patronage in the Hong Kong hotel industry. *International Journal of Hospitality Management*, 20(3), 277–297.
4. Fontenot, G., Henke, L., Carson, K., & Carson, P. P. (2007). Techniques for determining importance: Balancing scientific method and subjectivity. *Journal of Targeting, Measurement and Analysis for Marketing*, 15(3), 170–180.
5. Gustafsson, A., & Johnson, M. (2004). Determining Attribute Importance in a Service Satisfaction Model. *Journal of Service Research*, 7(2), 124–141.



6. Hu, N., Pavlou, P. A., & Zhang, J. (2007). Why do online product reviews have a J-shaped distribution? Overcoming biases in online word-of-mouth communication. *Marketing Science*, 19(8), 1037–1052.
7. Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM*, 52(10), 144–147.
8. Martin, D. W. (1995). An importance/performance analysis of service providers' perception of quality service in the hotel industry. *Journal of Hospitality & Leisure Marketing*, 3(1), 5–17.
9. Nadiri, H., & Hussain, K. (2005). Perceptions of service quality in North Cyprus hotels. *International Journal of Contemporary Hospitality Management*, 17(6), 469–480.
10. Pender, L., & Sharpley, R. (2004). *The management of tourism*. Sage.
11. Pokryshevskaya, E., & Antipov, E. (2014). A comparison of methods used to measure the importance of service attributes. *International Journal of Market Research*, 56(3), 283–296.
12. Rousseeuw, P. J., & Leroy, A. M. (2005). *Robust regression and outlier detection* (Vol. 589). John Wiley & Sons.
13. Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28(1), 180–182.