

## Survey Analytics of Star Hotels

Raj Aditya Kumar<sup>1\*</sup>, V.V Haragopal<sup>2</sup>

<sup>1</sup>Dept. of Computer Science, BITS, Pilani, Hyderabad, India

<sup>2</sup>Dept. of Mathematics, BITS, Pilani, Hyderabad, India

\*Corresponding Author: [rajaditya96@gmail.com](mailto:rajaditya96@gmail.com), Tel.: +91-9553337612

Available online at: [www.isroset.org](http://www.isroset.org)

Received:01/Oct/2018, Accepted:12/Oct/2018, Online: 31/Oct/2018

**Abstract**— The hospitality industry is growing not only in volume but in usage of technology as well. To keep pace with growing customer expectations, hotel industry needs to adapt quickly to ever changing customer preferences. This requires extensive use of analytics, driven by statistical methods, information technology and tools. Customer delight and the resulting high occupancy are the levers the hoteliers have, to stay competitive and improve profitability. In this article we discuss the preferences and views of the respondents about their likes and dislikes regarding the hospitality industry (star hotels). The primary data collected was analysed statistically with almost all attributes of preferences and the results are discussed in detail which will help the hospitality industry in making the right investment decisions. The approach is based on use of powerful multivariate statistical techniques like factor analysis to arrive at the conclusions. The statistical computations are analysed using SPSS (IBM-PASW) tool.

**Keywords**— Ordinal data, Eigen Values, Eigen Vectors, Principal Component Analysis, Factor Analysis

### I. INTRODUCTION

The hotel industry has seen a boom from late twentieth century. This is due to increasing globalization of business powered primarily by advances in Information technology. There has been an exponential increase in business travel both domestic and international. This has resulted in increasing demand for hotel.

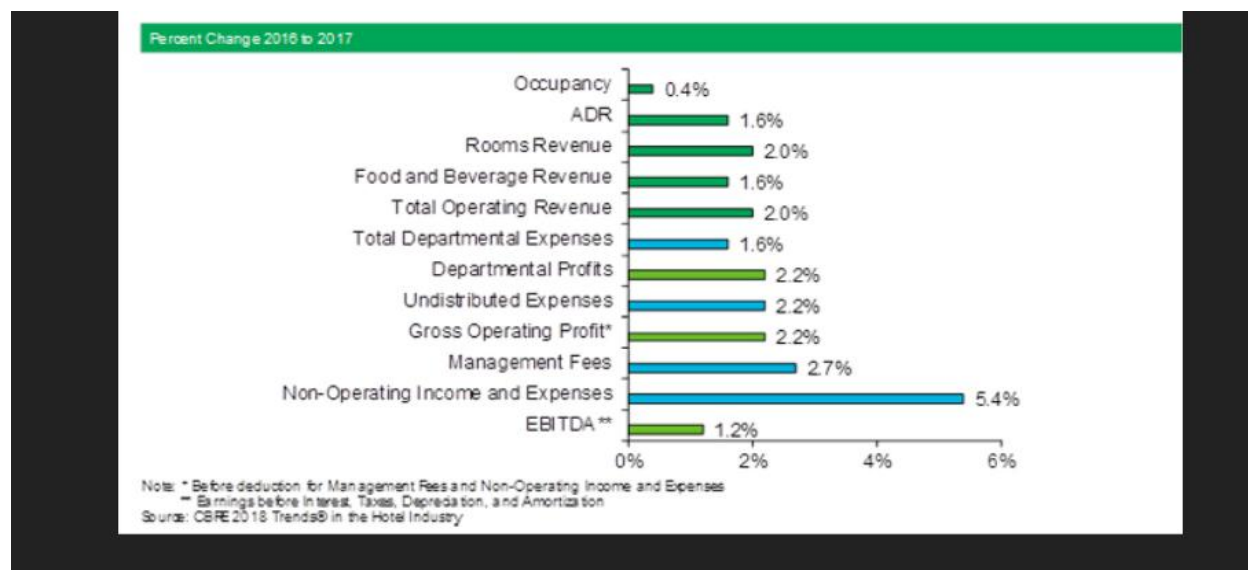


Figure 1 Growth trends in the Hotel Industry

**Figure 1** shows the growth trends in the hotel industry from 2016 to 2017. The occupancy has witnessed only 0.4% increase and the departmental profits increased only by 2.2%. It is really important that these numbers should be made more significant in line with the growing demands of the customers. Hence a primary survey was conducted which collected the customer feedback from their past experience or perceptions. Then the data is analyzed with multivariate statistical tool and promising conclusions are made. The statistical tool used to analyze the data here is factor analysis [1]. Factor analysis has been a popular technique adopted for statistical analysis. Brett Williams [2] gave a five step exploratory factor analysis protocol which provides novice researchers with starting reference point in developing clear decision pathways. N.Wu [3] proposed a factor analysis based anomaly detection algorithm which was applied to intrusion detection evaluation data. Aida Calvino [4] proposed a new method based on factor analysis to anonymize and protect datasets. S Ramana [5] used factor analysis to identify which regional features make traffic vulnerable to inclement weather.

In this paper, Section I contains an introduction stating the purpose of this study, brief history about the topic in study and previous research works. Section II contains the methodology used in this study. Section III contains the findings of this research. Section IV contains the

observations from the results which were obtained in Section III. Section V contains the conclusions of this study. The paper is completed with acknowledgement of the support we received and the references on this subject mentioned in Section VI and Section VII respectively.

**II. METHODOLOGY**

The below survey questionnaire shown from **Tables 1 to 3** was subjected on 100 respondents. This customer feedback data was collected and analysed. The data captured in the 5 point Likert scale was recorded and analysed by a multivariate technique. Factor analysis is especially popular in survey research, in which the responses to each question represent an outcome. As multiple questions often are related, underlying factors may influence subject responses. By analysing the correlations this technique involves data reduction, as it attempts to represent a set of variables by a smaller number.

Factor analysis attempts to discover the unexplained factors that influence the co-variation among multiple observations. These factors represent underlying concepts that cannot be adequately measured by a single variable.

**In Table 1, ranking data for hotel selection was captured as per the rating guidelines: 1 = LEAST IMPORTANT and 5 = MOST IMPORTANT**

**Table 1 Questionnaire to capture feedback data**

Past Experience	1	2	3	4	5
Budget	1	2	3	4	5
Discounts	1	2	3	4	5
Amenities	1	2	3	4	5
Hotel Reputation	1	2	3	4	5
Recommendations	1	2	3	4	5
Online reviews/Website	1	2	3	4	5
Advertisement	1	2	3	4	5
Location	1	2	3	4	5
Regular visitor here	1	2	3	4	5
Service	1	2	3	4	5
Food	1	2	3	4	5
Atmosphere	1	2	3	4	5

**In Table 2, ranking data for influence of good music was captured as per below guidelines.**

**1= Not at all    2= Slightly    3= Moderately    4= Very Much    5= Extremely**

**Table 2 Questionnaire to capture feedback data**

1	Entertainment - It lifts the mood and passes the time	<input type="checkbox"/>
2	Revival – It revitalizes and calms the mood	<input type="checkbox"/>

3	Strong sensation - It provides deep, thrilling emotional experiences	<input type="checkbox"/>
4	Diversion – It distracts the mind from unpleasant thoughts	<input type="checkbox"/>
5	Discharge – It releases emotions : purging and cleansing	<input type="checkbox"/>
6	Mental work – It encourages daydreaming, Nostalgia	<input type="checkbox"/>
7	Solace – It helps share emotions, experiences and connects to someone	<input type="checkbox"/>

In Table 3, ranking data was captured according to the following guidelines

1= Not at all    2= Slightly    3= Moderately    4= Very Much    5= Extremely

Table 3 Questionnaire to capture feedback data

Loud music influences my food and beverage selection.	1	2	3	4	5
Music tempo of the Luxury Hotel influences my dining speed.	1	2	3	4	5
Music of the Luxury Hotel decreases my perception of waiting time.	1	2	3	4	5
Music of the Luxury Hotel builds a positive "take-away" dining experience.	1	2	3	4	5
Music played in the Luxury Hotel suits time of the day.	1	2	3	4	5
The ambience scent of the Luxury Hotel is pleasant and welcoming, which improves my mood.	1	2	3	4	5
The ambience smell enhances my Hotel experience.	1	2	3	4	5
The efficiency of food service is a determining factor for me to visit the Hotel again.	1	2	3	4	5
A healthy menu with nutritive value helps me in making the choice to order a dish.	1	2	3	4	5

### III. STUDIES AND FINDINGS

We now introduce the concept of Factor analysis in this section.

#### Factor Analysis:

Many statistical methods are used to study the relationship between independent and dependent variables. Factor analysis is different and is used to study the patterns of relationship among many dependent variables with the goal of discovering something about the nature of the independent variables that affect them even though those independent variables were not measured directly. Thus answers obtained by factor analysis are necessarily more hypothetical and tentative than the case when independent variables are observed directly. The inferred independent variables are called factors.

Therefore, Factor analysis is a method for investigating how the variables of interest  $Y_1, Y_2, \dots, Y_k$  are linearly related to small number of unobservable factors  $F_1, F_2, \dots, F_m$  where  $m < k$ . Factor analysis measures the degree of linearity between the variables and factors.

The following is the mathematical formulation of factor analysis.

$$\begin{aligned}
 Y_1 &= B_{11}F_1 + B_{12}F_2 + B_{13}F_3 + \dots + B_{1m}F_m + e_1 \\
 Y_2 &= B_{21}F_1 + B_{22}F_2 + B_{23}F_3 + \dots + B_{2m}F_m + e_2 \\
 &\vdots \\
 Y_k &= B_{k1}F_1 + B_{k2}F_2 + B_{k3}F_3 + \dots + B_{km}F_m + e_k
 \end{aligned}$$

The error terms  $e_1, e_2, \dots, e_k$  serve to indicate that the hypothesized relationships are not exact.

The terms  $B_{ij}$  are called loadings. For instance  $B_{11}$  is the loading of  $Y_1$  on  $F_1$ . In simple terms  $B_{ij}$  is the degree of correlation between  $Y_i$  and  $F_j$  that is how much  $Y_i$  is related to  $F_j$ . For example, if  $B_{11} > B_{12}$  then  $Y_1$  is related more to  $F_1$  than to  $F_2$ .

Consider the term

$$Z_i = B_{i1}^2 + B_{i2}^2 + \dots + B_{im}^2$$

$Z_i$  is called the communality of  $Y_i$

$\frac{Z_i}{Var(Y_i)} \times 100$  is the percentage of variance explained by all

the factors  $F_1, F_2, \dots, F_m$ . Here  $Var(Y_i)$  is the variance of variable  $Y_i$ . Communality of a variable is the part of the variance explained by all the factors.

#### Big Picture of Communality

To make the concept of communality more clear as it plays a major role in this research conducted on hotel industry, let us consider the dimensionality point of view. Here there are k variables, in other words there are k dimensions. So each data point is a k-tuple. Now, it is intended to reduce the dimensionality of these data points from k to m (number of factors). But while reducing the dimensionality it might happen that some of the points overlap. Ideally this overlapping is not supposed to happen, but it usually

happens in practice. Hence the communality of the variable is the variation among the data points in the variable that is preserved even after reducing the dimensionality and for this reason, the communality of the variable is clearly less than the variance of the variable. If the communality of the variable is the highest then it can be concluded that the maximum variation in this variable is explained by all the factors and the variable is the most important because higher the communality, higher the percentage of variance preserved. Before performing the factor analysis, the communalities of the variables is 1. After the factor analysis the communalities are less than or equal to 1.

**Why Principal Component analysis in factor analysis?**

According to the mathematics of factor analysis

$$Var(Y_i) = Communality(Y_i) + Var(e_i)$$

$\forall i$  where  $e_i$  is the error term which specifies that the linear relationship between the variables and factors is not exact. Clearly different pairs of  $Communality(Y_i)$  and  $Var(e_i)$  satisfy the equation. The Communality of the variable is sum of squares of the loadings on the variable and so different values of loadings are possible for the variables considered.

Hence, Principal Component analysis is used to calculate the loadings which guarantees to bring the estimate of

communality of variable as close to the variance of the variable.

**Estimation of the number of factors and factor loadings**

The thumb rule used to find the number of factors is to take the covariance matrix and calculate its Eigen values. The number of Eigen values strictly greater than 1 is the number of factor components.

To estimate the factor loadings, we use Principal Component analysis.

After determining the number of factors we take the Eigen values corresponding to the factors. Let there be  $m$  factors and the Eigen values corresponding to the factors be  $v_1, v_2, \dots, v_m$ . Let the normalized Eigen vectors corresponding to the Eigen values be  $eig_1, eig_2, eig_3, \dots, eig_m$  respectively. The loadings of all variables on a particular factor  $F_i$  is given by  $\sqrt{v_i} \times eig_i$ .

**Research Findings**

Having dealt with the relevance of factor analysis, this technique has been applied on the Primary data collected on the attributes in Table 1.

The covariance matrix has been calculated based on the ratings of the customers. The following results shown from Tables 4 to 6 are obtained from the survey data

**Table 4 Total Variance Explained**

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.746	28.815	28.815	3.746	28.815	28.815
2	2.178	16.753	45.567	2.178	16.753	45.567
3	1.308	10.062	55.629	1.308	10.062	55.629
4	1.027	7.899	63.528	1.027	7.899	63.528
5	0.904	6.956	70.484			
6	0.757	5.824	76.308			
7	0.627	4.822	81.13			
8	0.586	4.509	85.639			
9	0.502	3.861	89.5			
10	0.466	3.586	93.086			
11	0.357	2.745	95.831			
12	0.292	2.25	98.081			
13	0.25	1.919	100			

**Table 5 Loading Matrix**

Attributes	Component			
	Influence of money	Influence of status	Influence of reviews and feedback	Influence of healthcare
Influence_Of_Past_Experience	0.484	-0.078	<b>0.539</b>	-0.128
Influence_Of_Budget	<b>0.502</b>	0.296	0.444	-0.482
Influence_Of_Discounts	<b>0.678</b>	0.33	0.313	-0.195
Influence_Of_Amenities	0.116	<b>0.589</b>	-0.137	0.006
Influence_Of_Hotel_Reputation	-0.045	<b>0.755</b>	-0.319	-0.071
Influence_Of_Recommendations	-0.368	0.387	<b>0.454</b>	-0.124
Influence_Of_Online_Reviews	0.225	-0.243	<b>0.723</b>	0.059
Influence_Of_Advertisement	0.158	0.073	<b>0.681</b>	0.452
Influence_Of_Location	0.088	<b>0.613</b>	-0.225	0.251
Influence_Of_Regular_Visitor_Here	0.349	-0.128	<b>0.634</b>	0.514
Influence_Of_Service	-0.376	<b>0.669</b>	0.068	0.154
Influence_Of_Food	-0.154	-0.458	-0.027	<b>0.698</b>
Influence_Of_Atmosphere	-0.012	-0.311	-0.164	<b>0.757</b>

**Table 6 Communalities**

Attributes	Initial	Final
Influence_Of_Past_Experience	1	0.548
Influence_Of_Budget	1	0.769
Influence_Of_Discounts	1	0.704
Influence_Of_Amenities	1	0.38
Influence_Of_Hotel_Reputation	1	0.679
Influence_Of_Recommendations	1	0.507
Influence_Of_Online_Reviews	1	0.636
Influence_Of_Advertisement	1	0.699
Influence_Of_Location	1	0.497
Influence_Of_Regular_Visitor_Here	1	0.805
Influence_Of_Service	1	0.617
Influence_Of_Food	1	0.722
Influence_Of_Atmosphere	1	0.697

**IV. OBSERVATIONS**

The following observations can be made for the attributes in **Table 1**, **Table 2** and **Table 3**:

- This bullet section completely talks about the factor analysis done on the attributes in **Table 1** only. The respondents’ ratings on these attributes are collected which shows their perceptions on how these attributes influence them to select a hotel. The following observation is made from the analysis of the data collected. In **Table 4** there are 4 Eigen values greater than 1, hence the 13 attributes in **Table 1** can be reduced to 4 factors. After performing the factor analysis, it can be seen in **Table 5** that “Influence of budget” and “Influence of discounts” are more strongly related to factor 1 than other factors with loadings of 0.502 and 0.678 respectively. These are shown in bold font. Hence factor 1 can be

named as “Influence of money” that is people do not want to spend a lot. This means questions about budget and discount are almost the same and are related to money. “Influence of Amenities”, “Influence of hotel reputation”, “Influence of location”, “Influence of service” are more strongly related to factor 2 than other factors with loadings of 0.589, 0.755, 0.615 and 0.669 respectively. Hence factor 2 can be named as “Influence of status” that is all these 4 attributes are similar and related to status because person with high status considers these attributes. “Influence of past experience”, “Influence of recommendations”, “Influence of online reviews”, “Influence of advertisements”, “Influence of regular visitor” are more strongly related to factor 3 than other factors with factor loadings of 0.539, 0.454, 0.723, 0.681, 0.634 respectively. Hence factor 3 can be named as “Influence of reviews and feedbacks” since all these attributes are similar and are more about reviews and feedbacks. “Influence of food”, “Influence of atmosphere” are strongly related to factor 4, hence factor 4 can be named as Influence of healthcare” since these attributes reflect a person with health consciousness. So next time while conducting a similar survey the 13 questions for 13 attributes in **Table 1** can be reduced to just 4 questions. This is particularly helpful because whenever a survey is conducted, new attributes might be considered other than the attributes shown in **Table 1**. Suppose 10 more attributes are considered, then there will be 23 attributes. But from the analysis done the 13 attributes considered now can be reduced to just 4 attributes. Hence there will be 14 attributes that would include the ten new attributes. From **Table 6** the communality of “Influence of Regular Visitor” is maximum (0.805), so the factor analysis best explains the variation in this attribute. That is all the 4 newly derived factors best explain this variable. The

Communalities of “Influence of advertisement, Influence of amenities, Influence of Location, Influence of online reviews” is much less. The Communality of “Influence of Food, Influence of Budget and Influence of discount” are also the 3rd, 2nd and 4th highest respectively. The conclusion from this observation can be referred to in Section V.

- Similar analysis was done on the attributes in **Table 2**. The respondents ratings on these attributes are collected which shows their perceptions on how good music influences them in the hotel. Since the analysis is similar, the results are not shown but the observation is as below:- The 7 attributes in **Table 2** can be reduced to 2 factors. After performing the factor analysis, it was observed that “music revival” and “music discharge” are strongly related to factor 2 and so factor 2 can be named as “mood calmer” because music revival and music discharge both are concerned with calming the mind and distracting the mind from unpleasant thoughts. All the other attributes are strongly related to factor 1 and hence factor 1 can be named as “mood lifter” since all the other attributes are concerned with boosting the mood. So during the next similar survey instead of considering these 7 attributes, just these 2 factors can be considered. The Communality of Music Diversion was observed to be the maximum which means that factor analysis best explains the variation in this attribute or we can say that all the factors (2 in this case) support this attribute. The Communality of Music Revival is 2nd largest which indicates that most of the customers even feel that music revitalizes and calms the mood. The conclusion from this observation can be referred to in Section V.
- Similar analysis was done on the attributes in **Table 3**. The respondents’ ratings on these attributes are collected which shows their perceptions on how the cleanliness enhances their hotel experience and influence of music on the dining experience. Since the analysis is similar, the results are not shown but the observation is as below:-

The 9 attributes in **Table 3** can be reduced to 2 factors. In this case the hidden independent factors are “Cleanliness and Quality of food service” and “Influence of music on the food”. The Communality of “The ambience smell enhances my hotel experience” is the maximum which indicates that both the factors support this attribute. Hence, for further research, just 2 attributes can be considered instead of 9 attributes. The conclusion from this observation can be referred in Section V.

## V. CONCLUSIONS

The following conclusions are based on the data collected and analyzed.

- This bullet section completely talks about the conclusions on the attributes in **Table 1**. From the factor analysis results obtained in section III and the first

observation in section IV it can be concluded that the majority of the customers who filled the survey feel that they will select a hotel if they are regular visitors of the hotel and if they are satisfied with the food and reasonable discounts are provided since the communalities of these attributes are high. The customers also feel that the advertisements, amenities, location and online reviews are not important since the communalities of these attributes are very less. So the hotel Communities can consider spending less time and money to advertise their hotel and providing amenities like swimming pool and instead can consider investing more on food (3rd highest communality), providing discounts, since customers consider their budget (2nd highest communality) for the selection of hotel.

- Similar type of conclusions are made on the attributes in **Table 2**. From the second observation in section IV it can be concluded that the majority of the customers believe that music distracts the mind from unpleasant thoughts since the Communality of Music Diversion is the maximum. The Communality of Music Revival is 2nd largest which indicates that most of the customers even feel that music revitalizes and calms the mood. Hence it can be concluded that good soothing music should be played in hotels while customers eat their food because their mood is calmed and they enjoy the food even more and will definitely consider visiting the same hotel again. Hence the hotel industry can consider investing on these attributes.
- Similarly for the attributes in **Table 3**. From the third observation in section IV it can be concluded that the majority of the customers feel that a good scent smell on entering the hotel makes them happy and feel pleasant since the communality of “The ambience smell enhances my hotel experience” is the maximum. This suggests that the hotel communities should maintain cleanliness and should employ labor to clean the complete hotel and should maintain air fresheners/fresh fragrance flowers so that good smell is spread throughout the hotel. So the hotel can consider investments accordingly. The smell of the cooked food should not spread in the hotel as that will lead to irritation in the customers. So the hotel can consider appropriate measures on this front. The communality of “Efficiency of food service” is the least which implies that most of the customers like self-service rather than being served food by any one. So the hotel communities can consider resorting to self service facilities.

On the whole, the Customer Feedback is the base for the growth of any business. Rather than subjective feedback, customers would prefer and be interested in the rating feedback. The hoteliers are interested in their customer satisfaction. Hence the customer satisfaction levels are analyzed using factor analysis tool

## VI. ACKNOWLEDGEMENT

The authors acknowledge Dr.Prarthana Kumar, Asst.Professor IPE Hyderabad for the use of her questionnaire

## VII. REFERENCES

- [1] Subhash Sharma, “*Applied Multivariate techniques*”, John Wiley & sons, Canada, pp. 90-128, 1996.
- [2] Brett Williams, Ted Brown, Andrys Onsman, “*Exploratory factor analysis: A five step guide for novices*”, Australasian Journal of Paramedicine, Vol. 8, Issue.3, pp.1-14, 2012.
- [3] N. Wu, J. Zhang, “*A factor analysis based anomaly detection*”, In IEEE systems, Man and cybernetics society information assurance workshop West Point, NY, USA, pp.1-10, 2003.
- [4] Aida Calvino, Palmira Aldeguer, Josep Domingo -Ferrer, “*Factor analysis for anonymization*”, In the 2017 IEEE International Conference on Data mining Workshops, New Orleans,LA,USA, pp.1-12, 2017.
- [5] S Ramana, S Sabitha, R Senthil Kumar, T Senthil Prakash “*Atmospheric Change on the Geographical Theme Finding Of Different Functions on Human Mobility*”, International Journal of Scientific Research in Computer Science and Engineering, Vol. 6, Issue.2, pp.134-151, 2018.

## AUTHORS' PROFILE

Raj Aditya K, Student in the 5th year of the 5 year Dual Degree ( M Sc Mathematics & B E (Hons) Computer Science) at BITS Pilani Hyderabad Campus. Raj Aditya's interests include application of statistical methods to develop new algorithms for Data mining and Machine Learning applications in Healthcare and Hospitality industries.



Dr. V. V Haragopal was a former Professor of Statistics, Head and Chairman Board of Statistics and Director CQM at Osmania University, Hyderabad. He has guided around 20 Ph.D. and 7 M. Phil students and published more than 100 papers in reputed international and national journals. He has delivered many invited talks across India. Presently he is working as Professor of Mathematics in Birla Institute of Technology and Science, Pilani at Hyderabad campus.

