



2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



ANALYSIS OF GUEST REVIEWS IN DIFFERENT STAR HOTELS AND TRIP TYPE AN APPLICATION IN SAO-PAULO

CH.V.MURALI KRISHNA, K.MANI CHAND, Y.TANUJA, K.SAI CHARITHA

NRI INSTITUTE OF TECHNOLOGY, Discipline of Computer Science Engineering JNTU-Kakinada University
Andhra Pradesh, India

Abstract :- Recommender Systems (RS) are proven to be very much beneficial on e-commerce sites by providing useful information to the customers in the decision-making process. The growth of computer mediated marketing has lead to increase interest in RS from past one decade Collaborative RS are the most popular type of these systems and they use ratings, i.e. the opinions of users on specific items, to determine neighbourhoods between users. Traditional RS like collaborative, content-based, knowledge-based, and hybrid systems use two dimensional ratings, i.e. the user and the item itself. But these systems omit much available information in some applications in which the contextual information plays a major role, for instance in the movies, restaurants or tourism domains. Contextual information has various forms like time, place, location, room, value, and etc. To acquire a dataset with the contextual information is not a simple task. For the purpose of analysis, we therefore consider a well-known e- tourism platform trip advisor (<http://www.tripadvisor.com>) which has many reviews given by different users with contextual information for various hotels across the city called sao-paulo The hotel industry in Sau-paulo is important, because number of fabulous events throughout the years, inviting visitors from all over the world as its guests. Hotels are primarily viewed as a service industry with intangible areas of guest experience and service. We focused on identifying the attributes that differentiate one hotel from another. The objective is to analyze the impact of contextual ratings on overall guest satisfaction & recommending potential segment with the help of multiple regression which is a part of Machine Learning. This analysis will help the hotel management(s). To concern the segments which are not highly co-related with overall rating. So that, they take the necessary steps to improve the concerns.

Keywords—contextual information, overall guest satisfaction, Multiple Regression, co-related

INTRODUCTION

Now a day's tourism Industry generates a significant portion of national income for many countries. Tourism industries are boost the country's economy and a large number of new job openings for the local people, and tourists spend money on a wide

range of services, including hotels, amusements transportation, food and medical Services, this way tourism yields an additional income, greatly supporting the country's. In South America, the tourism industry had an accelerated growth in

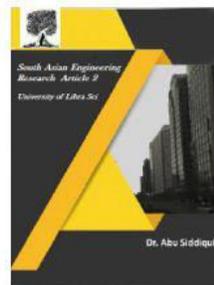


2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



2017. In this period a rate of 5.5 was reached according to MarketLine. This acceleration of industry is targetly due to economic development of Brazil with 61.7% of value of the tourism market. Saupaulo is part of Brazil. It is the largest city in Brazil. The hotel industry in Saupaulo is important, because number of fabulous events throughout the years, inviting visitors from all over the world as its guests. Hotels are primarily viewed as a service industry with intangible areas of guest experience and service levels. An analysis of hotel guest satisfaction ratings based on attributes such as location, sleep, quality, rooms etc., was performed. We focused on identifying the attributes that differentiate one hotel from another. For the purpose, we collect the data from one of the famous tourism domains which contain user opinion in the form of contextual ratings & overall guest satisfaction. The objective is to analyze the impact of contextual ratings on overall guest satisfaction recommending potential segment with the help of multiple regression which is a part of Machine Learning.

Results of this article will help the hotel management(s). To concern the segments which are not highly co-related with overall rating. So that they take the necessary steps to improve the concerns. The next section presents a literature review and the subsequent sections present the problem definition data and methodology and the results.

LITERATURE REVIEW

Recommender Systems (RS) use the opinions of a community to help individuals

in that community to identify the information or products most likely to be interesting for them. They have become fundamental applications in e-commerce and information filtering, providing suggestions that effectively prune large information spaces so that users are directed towards those products that meet their preferences. By combining ideas from user profiling, information filtering and machine learning, recommender systems have proven to be elective at delivering the user proactively information. These RS are basically classified into content based, collaborative, and knowledge-based systems based upon their approach to rating estimation. Any of these recommendation techniques can be combined to form hybrid recommender systems. These traditional RS have been focussing on recommending the most relevant and appropriate items to users. Although these traditional systems have performed reasonably well in several applications but however in many other applications, such as time and location-based recommendations, it may not be sufficient to consider only customers and items. This analysis is all about multi-dimensional RS that include the contextual information such as value, room, location etc in to the recommendation process for better performance. We decided to extract data from tripadvisor (<http://www.tripadvisor.com>), which is a very popular website for e-tourism. In that we extracted the reviews given by different users for various hotels which include contextual information. The contextual

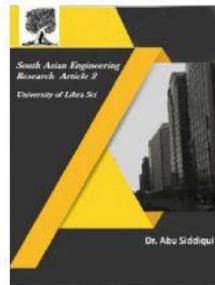


2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



segments which are expressed in terms of ratings are Value, Rooms, Location, Cleanliness, Service, and Business. The overall rating of the review was also studied. All these ratings are expressed on a 5-point Like scale i.e. 1 - 5. The extracted data also contains several other contextual segments which are expressed in terms of predefined options such as companion of the trip, purpose of the visit etc.

PROBLEM DEFINITION

There are many aspects to measure hotels performances in terms of its Customers experiences and customer satisfaction. Customer feedback in terms of ratings Collected from hotels located in Sao Paulo from 2018 to 2019 using trip advisor. The objective is to analyze the impact of contextual ratings such as maintenance, Service, value for money, cleanliness and overall satisfaction on overall guest satisfaction .Another objective is to analyze the customer ratings based on trip such as family, couple, Business, singles using statistical analysis, for each hotel in SaoPaulo.

SOLUTION TO PROBLEM DEFINITION

A comparative study has been done on different star hotels data in Sao Paulo from a Trip Advisor, and analyses the individual context rating such as sleep, location, rooms, service, maintenance, value on overall rating given by the customer by using multiple regression. The Objective of this study was to understand the satisfaction levels of the hotel customers. The result was obtained each of the hotels in Sao Paulo by using

quantitative analysis to better understand what customers like and dislike about each hotel.

DATA AND METHODOLOGY

The data used in this study were retrieved from Trip Advisor using Webscraping. TripAdvisor has a mechanism that enables users to input reviews for their hotel stays and share experiences, posting classifications from 1 (very unsatisfied/terrible) to 5 (very satisfied/excellent).The methodology used in this study was explained below with the help of process diagram.

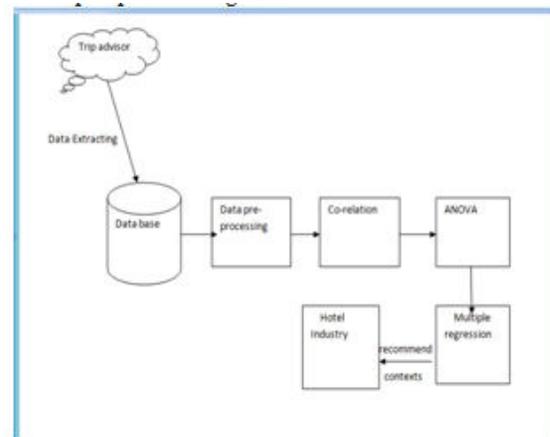


Fig1:-Process Diagram

Web Data Extraction:-

To extract information from any web source, we need an open source browser to connect to the server, and a parser to parse the source tree of the website. The connection should be, stable all the time, while extracting data from the website. Any break downs in the connection result in abnormal data rather than actual data. We decided to use standard libraries in python to achieve data.

Urllib.request:-

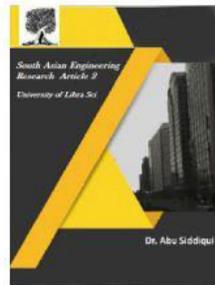


2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



It is a Python module for fetching URLs (Uniform Resource Locators). It offers a very simple interface, in the form of the `urlopen` function.

This is capable of fetching URLs using a variety of different protocols. `Urllib.request` supports fetching URLs for many “URL schemes” (identified by the string before the in URL - for example “ftp” is the URL scheme of “ftp://python.org/”) using their associated network protocols (e.g. FTP, HTTP). It also offers a slightly more complex interface for handling common situations - like basic authentication, cookies, proxies and so on. These are provided by objects called handlers and openers.

For straightforward situations `urlopen` is very easy to use. But as soon as you encounter errors or non-trivial cases when opening HTTP URLs, you will need some understanding of the HyperText Transfer Protocol. The most comprehensive and authoritative reference to HTTP . This is a technical document and not intended to be easy to read. This HOWTO aims to illustrate using `urllib`, with enough detail about HTTP to help you through. It is not intended to replace the `urllib.request` docs, but is supplementary to them. Many uses of `urllib` will be that simple (note that instead of an ‘http:’ URL we could have used a URL starting with ‘ftp:’, ‘file:’, etc.). However, it’s the purpose of this tutorial to explain the more complicated cases, concentrating on HTTP.

HTTP is based on requests and responses - the client makes requests and servers send

responses. `Urllib.request` mirrors this with a request object which represents the HTTP request you are making. In its simplest form you create a Request object that specifies the URL you want to fetch. Calling `urlopen` with this Request object returns a response object for the URL requested. This response is a file-like object, which means you can for example call `read()` on the response.

BEAUTIFUL SOUP:-

Beautiful Soup is a Python library aimed at helping programmers who are trying to scrape data from websites. To use beautiful soup, you need to install it: `$ pip install beautifulsoup4`. Beautiful Soup also relies on a parser, the default is `lxml`. You may already have it, but you should check (open IDLE and attempt to import `lxml`). If not, do: `$ pip install lxml` or `$ apt-get install python-lxml`

Import library

Now you have to pass something to BeautifulSoup to create a soup object.

BeautifulSoup does not fetch the web page for you, you have to do that yourself.

That’s why I use `urllib` in combination with the BeautifulSoup library.

There are some different filters you can use with the search API.

Below I will show you some examples on how you can pass those filters into methods such as `find_all`

The `find_all` method is one of the most common methods in BeautifulSoup.

It looks through a tag’s descendants and retrieves all descendants that match your filters.

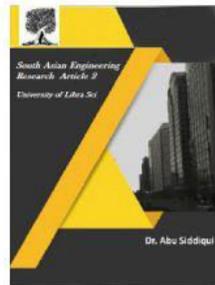


2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



If you pass in a regular expression object, BeautifulSoup will filter against that regular expression using its match() method.

SELENIUM: -

Selenium is a popular open-source web-based automation tool. selenium is a free(open-source) automated testing framework used to validate web applications across different browsers and platforms. you can use multiple programming languages like java, c#, python etc to create. selenium test scripts. Testing done using the selenium tool is usually referred to as selenium testing.

Simon Stewart created WebDriver circa 2006 when browsers and web applications were becoming more powerful and more restrictive with javascript programs like selenium co. It was the first cross-platform testing framework that could control the browser from the OS level.

SELENIUM WEBDRIVER:-

- To use a certain programming language in designing your test case.
- To test applications that are rich in AJAX-based functionalities.
- To execute tests on the HtmlUnit browser.
- To create customized test results.

The WebDriver proves itself to be better than both selenium IDE and selenium RC in many aspects. It implements a more modern and stable approach in automating the browser's actions. WebDriver, unlike Selenium RC, does not rely on JavaScript for Automation. It controls the browsers by directly communicating with it.

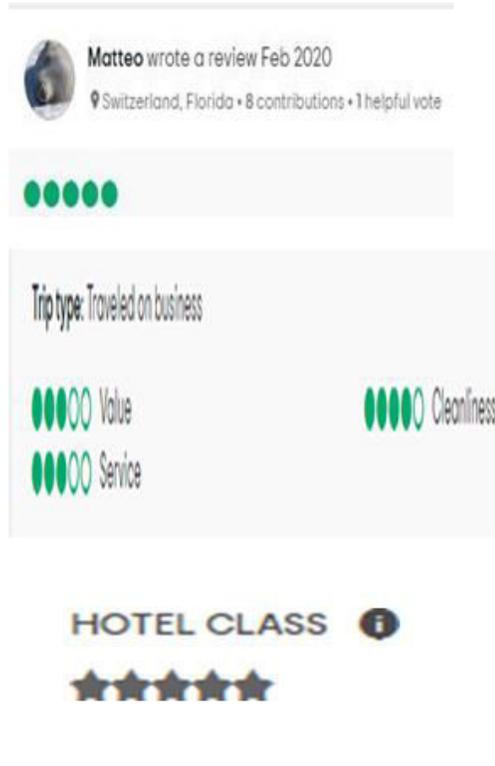


Fig2:-example sample data we extract

ANALYZING DATA:-

DATA PREPROCESSING:-

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. Data preprocessing prepares raw data for further processing.

Data Cleaning:-

In our study we mainly concentrate ratings for analysis purpose for this reason our data does not have noisy, data has only missing values problem. Data cleaning is through processes such as filling in missing

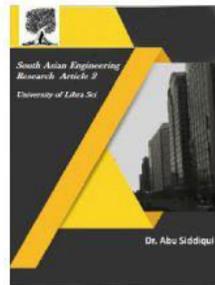


2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



values in the data. Missing values are a common occurrence, and you need to have a strategy for treating them. A missing value can signify a number of different things in your data. Perhaps the data was not available or not applicable or the event did not happen. It could be that the person who entered the data did not know the right value, or missed filling in. Data mining methods vary in the way they treat missing values. Typically, they ignore the missing values, or exclude any records containing missing values, or replace missing values with the mean, or infer missing values from existing values.

we specifically look at dealing with missing data, there are several techniques that can be used. Choosing the right technique is a choice that depends on the problem domain — the data’s domain (sales data? CRM data? ...) and our goal for the data mining process

1.Ignore the datarow:-

This is usually done when the class label is missing (assuming your data mining goal is classification), or many attributes are missing from the row (not just one). However, you’ll obviously get poor performance if the percentage of such rows is high. For example, let’s say we have a database of students enrolment data (age, SAT score, state of residence, etc.) and a column classifying their success in college to “Low”, “Medium” and “High”. Let’s say our goal is to build a model predicting a student’s success in college. Data rows who are missing the success column are not useful in predicting success so they could

very well be ignored and removed before running the algorithm.

2. Use a global constant to fill in for missing values:-

Decide on a new global constant value, like “unknown“, “N/A” or minus infinity, that will be used to fill all the missing values. This technique is used because sometimes it just doesn’t make sense to try and predict the missing value. For example, let’s look at the students enrolment database again. Assuming the state of residence attribute data is missing for some students. Filling it up with some state doesn’t really makes sense as opposed to using something like “N/A”.

3.Use attribute mean:-

Replace missing values of an attribute with the mean (or median if its discrete) value for that attribute in the database. Forexample, in a database of US family incomes, if the average income of a US family is X you can use that value to replace missing income values.

4. Use attribute mean for all samples belonging to the same class:-

Instead of using the mean (or median) of a certain attribute calculated by looking at all the rows in a database, we can limit the calculations to the relevant class to make the value more relevant to the row we’re looking at. Let’s say you have a cars pricing database that, among other things, classifies cars to “Luxury” and “Low budget” and you’re dealing with missing values in the cost field. Replacing missing cost of a luxury car with the average cost of all luxury cars is probably more accurate than the

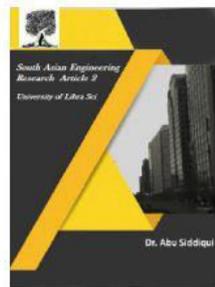


2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



value you'd get if you factor in the low budget cars.

5. Use a data mining algorithm to predict the most probable value:-

The value can be determined using regression, inference based tools using Bayesian formalism, decision trees, clustering algorithms (K- Mean\Median etc.). For example, we could use clustering algorithms to create clusters of rows which will then be used for calculating an attribute mean or median as specified in technique #3. Another example could be using a decision tree to try and predict the probable value in the missing attribute, according to other attributes in the data. I'd suggest looking into regression and decision trees first (ID3 tree generation) as they're relatively easy and there are plenty of examples on the net...

Compared to all above techniques mean is the efficient technique for our data.

Correlation:-

Variables within a dataset can be related for lots of reasons. For example:-

One variable could cause or depend on the values of another variable. One variable could be lightly associated with another variable.

Two variables could depend on a third unknown variable.

It can be useful in data analysis and modeling to better understand the relationships between variables. The statistical relationship between two variables is referred to as their correlation A correlation could be positive, meaning both

variables move in the same direction, or negative, meaning that when one variable's value increases, the other variables' values decrease. Correlation can also be neutral or zero, meaning that the variables are unrelated. Positive Correlation: both variables change in the same direction.

Neutral Correlation: No relationship in the change of the variables.

Negative Correlation: variables change in opposite directions. The performance of some algorithms can deteriorate if two or more variables are tightly related, called multicollinearity. An example is linear regression, where one of the offending correlated variables should be removed in order to improve the skill of the model. We may also be interested in the correlation between input variables with the output variable in order provide

insight into which variables may or may not be relevant as input for developing a model.

The structure of the relationship may be known, e.g. it may be linear, or we may have no idea whether a relationship exists between two variables or what structure it may take. Depending what is known about the relationship and the distribution of the variables, different correlation scores can be calculated.

The Pearson correlation coefficient (named for Karl Pearson) can be used to summarize the strength of the linear relationship between two data samples. The Pearson's correlation coefficient is calculated as the covariance of the two variables divided by the product of the standard deviation of each data sample. It is the

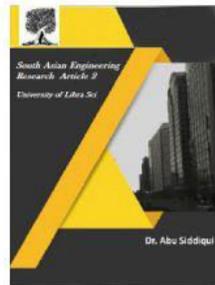


2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



normalization of the covariance between the two variables to give an interpretable score. Python is a great language for doing data analysis, primarily because of the fantastic ecosystem of data-centric python packages. Pandas is one of those packages and makes importing and analyzing data much easier. Pandas dataframe.corr() is used to find the pairwise correlation of all columns in the dataframe. Any na values are automatically excluded. For any non-numeric data type columns in the dataframe it is ignored. Use corr() function to find the correlation among the columns in the dataframe using 'Pearson' method. The output data frame can be interpreted as for any cell, row variable correlation with the column variable is the value of the cell. As mentioned earlier, that the correlation of a variable with itself is 1. For that reason all the diagonal values are 1.00.

ANOVA:-

ANOVA is a quick, easy way to rule out unneeded variables that contribute little to the explanation of a dependent variable. It is accessible and applicable to people outside of the statistics field.

This instructable will assume no prior knowledge in R and will give basic software commands that may be trivial to an experienced user. If you are familiar with R I suggest skipping to Step 4, and proceeding with a known dataset already in R.

R is a free, open source, and ubiquitous in the statistics field. R has all-text commands written in the computer language S. It is helpful, but by no means necessary, to have an elementary understanding of text based

computer languages. If you can not stand working with text-based software I suggest that you try the statistics software JMP. The case with multiple independent variables attribute1,attribute2 to attribute n is a simple change.
name=aov(overallrating~attribute1+attribute 2+.....attribute(n))

MULTIPLE REGRESSION:-

Multiple Linear Regression attempts to model the relationship between two or more features and a response by fitting a linear equation to observed data. The steps to perform multiple linear Regression are almost similar to that of simple linear Regression. The Difference Lies in the evaluation. We can use it to find out which factor has the highest impact on the predicted output and now different variable relate to each other.

$$\text{Here: } Y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + \dots + b_n * x_n$$

RESULTS

After Scraping:

total number of user ratings	total size of total dataframes	value	location	date	url		
Scoutor 1	Travelled to	January 2 Family Bu	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_1	
Scoutor 2	Travelled to	December Family Bu	5	4	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_2
Scoutor 3	Travelled to	October 2 Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_3
Scoutor 4	Travelled to	November Family Bu	5	5	5	5	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_4
Scoutor 5	Travelled to	November Family Bu	4	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_5
Scoutor 6	Travelled to	November Family Bu	4	5	5	5	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_6
Scoutor 7	Travelled to	October Family Bu	4	4	5	5	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_7
Scoutor 8	Travelled to	November Family Bu	5	5	5	5	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_8
Scoutor 9	Travelled to	October Family Bu	4	5	5	5	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_9
Scoutor 10	Travelled to	October Family Bu	5	5	5	5	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_10
Scoutor 11	Travelled to	February Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_11
Scoutor 12	Travelled to	October Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_12
Scoutor 13	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_13
Scoutor 14	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_14
Scoutor 15	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_15
Scoutor 16	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_16
Scoutor 17	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_17
Scoutor 18	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_18
Scoutor 19	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_19
Scoutor 20	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_20
Scoutor 21	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_21
Scoutor 22	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_22
Scoutor 23	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_23
Scoutor 24	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_24
Scoutor 25	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_25
Scoutor 26	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_26
Scoutor 27	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_27
Scoutor 28	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_28
Scoutor 29	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_29
Scoutor 30	Travelled to	September Family Bu	10	10	10	10	https://www.tripadvisor.in/Hotel_Review-g18862-G48883-Reviews-Scoutor_30

Fig3:-Table1

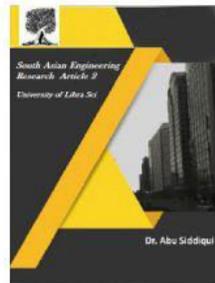


2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



After Preprocessing:-

over_all_rating	trip_type	hotel_class	cleanliness	service	value	location
4	Travelled on business	2	1	2	1	1
4	Travelled as a couple	2	5	4	1	1
4	Travelled on business	2	1	2	1	1
4	Travelled on business	2	5	5	5	5
4	Travelled on business	2	1	4	1	1
4 na		2	1	2	1	1
4 na		2	1	2	1	1
4 na		2	1	2	1	1
4 na		2	1	2	1	1
4 na		2	1	2	1	1
4 na		2	5	5	5	1
4	Travelled on business	2	5	4	5	4
4	Travelled on business	2	4	4	4	4
4	Travelled with family	2	1	2	1	1
4	Travelled on business	2	4	3	3	3
4 na		2	1	2	1	1
4 na		2	1	2	1	1
4 na		2	1	2	1	1
4 na		2	1	2	1	1
4	Travelled on business	2	1	4	4	1
4	Travelled on business	2	5	5	1	5
4	Travelled on business	2	5	4	5	4
4	Travelled with friends	2	1	5	4	1

Fig4:-Table2

Correlation results:-

	over_all_rating	hotel_class	cleanliness	service	value	location
over_all_rating	1.000000	0.674031	-0.060964	-0.000345	-0.072969	-0.121802
hotel_class	0.674031	1.000000	-0.103410	-0.035477	-0.109390	-0.113518
cleanliness	-0.060964	-0.103410	1.000000	0.696118	0.638960	0.680141
service	-0.000345	-0.035477	0.696118	1.000000	0.758850	0.704164
value	-0.072969	-0.109390	0.638960	0.758850	1.000000	0.670391
location	-0.121802	-0.113518	0.680141	0.704164	0.670391	1.000000

Fig5:-Table3

```

> summary(aov(over_all_rating~trip_type,df))
      Df Sum Sq Mean Sq F value Pr(>F)
trip_type  5  2.34  0.4684   3.71 0.00244 **
Residuals 1290 162.89  0.1263

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

```

Fig6:-Table4

Multiple regression result:-

```

Coefficients:
(Intercept)          3.679909  0.040342  91.217 < 2e-16 ***
trip_type Travelled on business -0.007693  0.035389  -0.217  0.82794
trip_type Travelled solo      -0.083901  0.053819  -1.559  0.11925
trip_type Travelled with family -0.073080  0.051406  -1.422  0.15538
trip_type Travelled with friends -0.145791  0.055548  -2.625  0.00878 **
trip_type na                 -0.013541  0.033614  -0.403  0.68713
cleanliness           0.018837  0.008803   2.140  0.03254 *
value                 0.008879  0.009454   0.939  0.34786
location             -0.036251  0.008992  -4.031  5.87e-05 ***
hotel_class           0.118725  0.005109  23.239 < 2e-16 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2969 on 1286 degrees of freedom
Multiple R-squared:  0.3139, Adjusted R-squared:  0.3091
F-statistic: 65.38 on 9 and 1286 DF, p-value: < 2.2e-16

```

Fig7:-Table5

CONCLUSION

Tourism industry is growing at a rapid rate across the world which will become the largest industry in the world generating about 13.9 percentage of GDP and employment . The customers share their views and ideas about their visited hotels through web application platforms that have proved very useful for new customers. So, customer satisfaction has proved to be a very crucial factor for management to sustain in the market. Multi dimensional RS are becoming very useful in these e-tourism web applications since they can include all the contextual information that the customer provided in his review about his own experiences regarding a particular destination. Thus, they can include the review information in terms of additional dimensions. The review rating expressed by a customer is an indication of his/her opinion and also an indication for satisfaction/dissatisfaction. But this overall rating is a We are mainly focus on



2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



qualitative analysis that means we analyzing the combination of various other aspects which are expressed as multiple dimensions in the given scenario. Therefore, the association between these multiple dimensions and the overall rating is a crucial factor for identifying the responsible segments for customer opinion.

This association can be identified by using a statistical approach called regression modeling individual attributes on overall guest ratings based on hotel class and triptype . After analyzing the data produced result was recommended to the hotel industry.

References

- [1] Cobra: Java HTML Parser. <http://lobobrowser.org/cobra/java-html-parser.jsp>.
- [2] Maryam Ramezani, Andriy Shepitsen, Runa Bhaumik, Robin Burke, and Bamshad Mobasher. Using Semantic Cues for Contextual Recommendation. Technical report, Depaul University, School of Computer Science, Telecommunicaitons and Information Systems.
- [3] M. Gorgoglione, C. Palmisano, and A. Tuzhilin. Personalization in Context: Does Context Matter When Building Personalized Customer Models? In Proceedings of the Sixth International Conference on Data Mining(ICDM '06), pages 110. IEEE, 2006.
- [4] Soane Abbar, Mokrane Bouzeghoub, and Stéphane Lope. ContextAware Recommender Systems: A ServiceOriented Approach. pages 16. ACM, 2009.
- [5] Gediminas Adomavicius, Alexander Tuzhilin, Ramesh Sankaranarayanan, and Shahana Sen. Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach. ACM Transactions on Information Systems, 23(1):103145, 2005.
- [6] Gediminas Adomavicius and Alexander Tuzhilin. Context-Aware Recommender Systems. In 2nd ACM International Conference on Recommender Systems(RecSys '08), page 235. ACM, 2008
- [7] Cosimo Palmisano, Alexander Tuzhilin, and Michele Gorgoglione. Using Context to Improve Predictive Modeling of Customers in Personalization Applications. IEEE Transactions on Knowledge and Data Egnineering, 20(11):15351549, 2008vvvvv
- [8] Francesco Ricci, Lior Rokach, Brancha Shapira, and Paul B. Kantor. Recommender Systems in E- Commerce: Recommender Systems Handbook. Springer, May, 2010.
- [9] Gediminas Adomavicius and YoungOk Kwon. New Recommendation Techniques for Multicriteria Rating Systems. In IEEE Intelligent Systems, pages 4855. IEEE, June 2007
- [10] The Lobo Project. <http://lobobrowser.org/java-browser.jsp>.
- [11] https://www.researchgate.net/publication/319399409_Customer_Satisfaction_in_Portuguese_Hotels_Evidence_for_Different_Regions_and_Hotel_Segments
- [12] https://www.researchgate.net/publication/319399409_Customer_Satisfaction_in_Portuguese_Hotels_Evidence_for_Different_Regions_and_Hotel_Segments

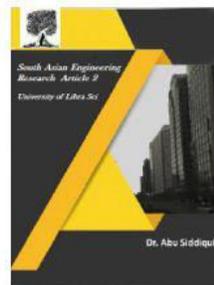


2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



n/2918328 00_Analysis_of_Hotel_Gu
est_Satisfaction_Ratings_and_Reviews_An_
Applicatio n_in_Singapore