Impact Of Attribute Ratings On Overall Hotel Ratings: A Sentiment Analysis Of Indian Five-Star Hotels Online Reviews

Anil Kumar[1], Subhabaha Pal[2], Kamakshi Mehta[3]

[1]Research Scholar Manipal University Jaipur, Email: mydream.mba2010@gmail.com
[2]Assistant Professor Manipal University Jaipur, Email: subhabaha.pal@jaipur.manipal.edu
[3]Head and Associate Professor TAPMI School of Business, Manipal University Jaipur,
Email: kamashi.mehta@jaipur.manipal.edu

ABSTRACT
The present study scrutinizes the behaviour, attitude and sentiments of consumers who have visits to five-star hotels across major eight cities in India "i.e. Delhi, Gurugram, Chandigarh, Kolkata, Mumbai, Bangalore, Jaipur and Chennai". The sentiments are articulated through online reviews and ratings on the renowned 'Tripadvisor' website. Hotel reservation web portal utilize these ratings and online reviews to guide the customer for making informed decisions for hotel selection and provide insights about the hotel services, facilities and amenities. However, many travelers' not had enough patience or time to engage with each review and rating.
This study aim to assess the effect of online hotel reviews furnishes enough information pertaining to customer behaviour, attitude and sentiment across different parameters. Further this study endeavors to determine various factors influence the overall rating of the five star-hotels
Additionally, ascertain that value rating have significant influence on overall rating of the establishment.
The finding of this study contributes significantly for both academic discourse and hotel manager practices implementation. They furnish valuable insight to hotel managers, enabling them to formulate informed strategies and policies aligned to an enhanced comprehension understanding of consumer behaviour and attitude towards their hotel.

Key words: Sentiment Analysis, Online Reviews, Data Scraping, Social Media, Ratings.

1. INTRODUCTION
In contemporary times, users are interested to adopt convenient methods to innovate for gaining the information and knowledge through Internet. This trend has become popular to search and find a suitable and convenient place for stay other during their journey. Consequently, the hotel industry has significant under transformation and uses the websites for advertising, brand building and image improvement for which large number of websites are available. This study is conducted because in India very few researches are conducted on five-star hotels.

1.1 Online Reviews and Ratings
User-generated contents and data have emerged as a pivotal source of information particularly within the hotel industry, where consumers want real, unbiased and unregulated information. The consumer's generated online reviews are such source of information that customers finds reliable and trustworthy. Analyzing these online text reviews are helping to understand the behaviour, attitude and experience of the customers in the hotels considered in the various studies in the literature. Some renowned websites in India are MakemyTrip, BookmyTrip, GoIbibi, Tripadvisor, Yatra, Agoda, etc. Through these websites users express their expectations for a particular hotel when anyone visit or would not match with their budget.

1.2 Indian Five Star hotels
This study focused on extracted reviews from the TripAdvisor pertaining to five-star hotels across eight different cities of India namely Delhi, Gurugram, Chandigarh, Kolkata, Mumbai, Bangalore, Jaipur and Chenna. Indian hotel industry is one of the fastest growing industry in the India (17th Rank in world 2022, 3rd Rank in Asia Pacific 2021) as per 'Tourism Report 2023 and hotel industry contributing to GDP $40 billion in 2022 and is expected to reach $68 billion by 2027 and $1 trillion expected to 2047 (Economics Times).

Five-Star hotels are selected for this study due to their prominence in this industry landscape, having significant portion out of 100,980 approved rooms (One Star – Five Star) 64002 rooms allocated in five star hotels and occupancy percentage of hotels in India 65% in 2023 as per tourism report 2023. Utilizing technical tools for reviews analysis and leveraging social media platform to attract more and more customers to enhance occupancy rates.

1.3 Sentiment Analysis

Sentiment Analysis emerges as best techniques to extract the opinion unstructured data which generated by the consumers on the various online platforms that also very important factor in business decision making and business intelligence. The current study employs sentiment and polarity analysis to gain insight into customers' opinion.

Now in contemporary hospitality landscape, hotels need quick feedback from its customers that will help for addition and modification of products and services. So the hotel must diligently observe the feedback given by the customers in the form of reviews and rating on various websites. The reviews given by the customers on the various websites can be analysis with the help of various technical tools i.e R-Studio, Python & Panda etc. these help to find sentiment and polarity analysis. For a hotel business various aspect like Cleanliness, Food, Staff, Hospitality, Room neatness, etc, perform a significant role in recommendation system. The consumer's sentiment with respect to a particular hotel depends on the facilities they receives such as price, location, environment etc., and service provided by hotel. The Sentiment can be expressed in the form of ten parameters. In this study reviews are extracted from the 'Tripadvisor' website related to the five star hotels of the different eight cities of India "Delhi, Gurugram, Chandigarh, Kolkata, Mumbai, Bangalore, Jaipur and Chennai" through which customers express their experience.

2. LITERATURE REVIEW

The Consumer reviews provide a cost effective way to monitor the views and feedback given by the customer and can gain the competitive advantage over the competitor(Stuart, Mednick, & Bockman, n.d.). The sentiment analysis done through the modern methods provide better accuracy as compared through the survey methods (Sharad, Ashok, Dattatray, & Bhagwat, 2018).Trend of online heterogeneous data is increased on the various websites and machine learning approach can be used to analyses (Sanwal & Kukreja, 2019).Aspect based sentiment analysis on the various factors provide clarity in the results (Rybakov & Malafeev, n.d.). The facilities and surrounding of the hotel not deliver a positive impact on the sentiments of the customer(Raul & Silva, 2019). The sentiment of the consumers will vary according to the various aspects of the hotel like service, food, location etc. (Priyantina, 2019). The aspect based sentiment analysis in which consumer sentiments are influenced by the various aspects of the hotel and cleanliness is the most discussed aspect (Pratama, 2022). Hotel sentiment is very important in analysing the role of reviews in exploring the many kind of information regarding the hotel recommendation and quality of service (Mrkopoulos, Mikros and Iliadi, 2015). The sentiment analysis play an import role to select the hotel according to their need and also significant to know the area of improvement by using the machine learning(Kanade, 2019).The sentiment oriented model used to describe the reviews in the form of positive, average and negative and that will help the hotel to provide the recommendation to the customers for selecting the hotel (Elango & Narayanan, 2002). Three-dimensions grid provide the results in the form of risk and utilities by compare with no trade-off and trade-off to explain the hotel attributes that helps the customer to take decision (Dolnicar, 2003).The online reviews expressed by customer on the Tripadvisor in which customer rating are not only sufficient the reviews sentiment analysis play an important role for customers and hotel managers (Dina, 2020). Aspect based sentiment analysis helps the organisation to focus on the particular area on which improvement is required as well as sentiment score is much more different from the rating given on the website (Akhtar, Zubair, Kumar, & Ahmad, 2017).

3. Objectives

- Finding the attributes rating that has highest impact on overall rating of hotel.
- To analyze behaviour, attitude and sentiment of customers.
- Finding the polarity of online reviews.

4. Hypotheses

H1: Overall Hotel rating positively influenced by Rooms rating.
H2: Overall Hotel rating positively influenced by Cleanliness rating.
H3: Overall Hotel rating positively influenced by Service rating.
H4: Overall Hotel rating positively influenced by Value rating.

5. RESEARCH METHODOLOGY
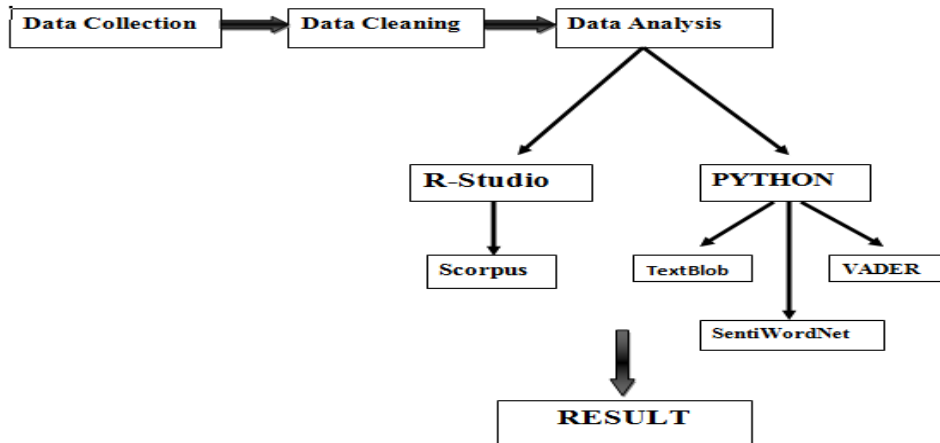
5.1 Data Collection

TripAdvisor website was selected for data collection because TripAdvisor is one of the most visited website in nation and contains data about allmost of five star hotels in the form of reviews which given by the customers related to the hotel quality,staff and services. In this study 3326 reviews are scraped with the help of "ProWebScraper" of the Indian five star hotels of different major eight cities of India in the form of "CSV" file. It takes as the input the of hotel's webpage on the TripAdvisor website and then crawls the ten reviews ata time. It creates one file contains the reviews text and other matadata. The Metadata includes the following:

- Review Title
- Review Date
- Name of Guest
- Review Text
- Hotel url

For finding the impact of hotel attribute rating like i.e Location, Cleanliness, Service, Value rating impact on overall rating of the hotel rating of 100 five star hotels from major eight cities are collected.

5.2 Data Cleaning

The next step through which reviews csv file imported to the 'R-studio' and various packages are loaded from the library like text mining (tm),NLP, dplyr etc. than corpus is build. Through R-Studio only reviews text is selected and reviews contain many unuseful containt like punctuations, numbers, stopwords these are removed by using the corpus. After this 'Cleanset is developed to remove the most repeating common words and whitespace.

5.3 Data analysis

To find out the better results in our study we use to technical tool R-Studio & Python. In further process by using the 'cleanset' tdm- Term Document Matrix is developed to find out the maximum repeating words throughout the reviews and that can be visualized through word cloud in that maximum repeating words highlighting ex. Hotel, food, staff, stay, room, service it means customers are talking about in the reviews. In the reviews most repeated words are ex. amazing, excellent; thank, best, great, special, nice and comfortable etc. that showing customers use the positive words and appreciate the hotel service, staff and rooms as shown in figure.1

Figure 1. Most repeating words in the reviews

Corpus Based Sentiment Analysis (R-Studio)

In the analysis process of reviews sentiment score is finding out of the Indian five star hotels on different ten factors by using package 'sentimentr' that will show the score of the reviews on the different ten parameter as shown in the figure.2
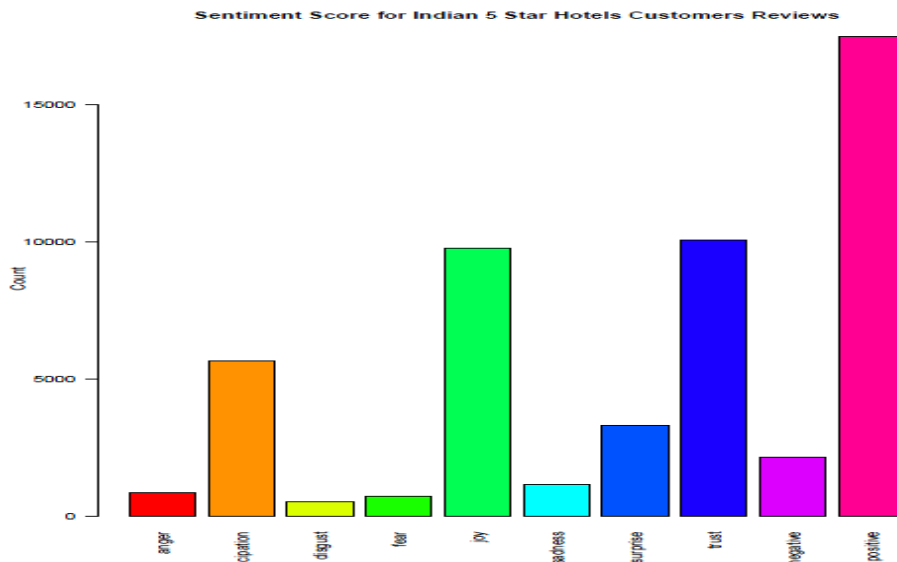
Figure 2. Sentiment score of the Reviews with its Visualization

Bar graph and score clearly explains that the reviews those are presented through the 'TripAdvisor' having liking and favouring attitude towards the Indian five star hotels service quality, physical attributes and amenities as per the graph and score there are very less number of negative reviews if compare with the positive response. The sentiment analysis done through the R-Studio anger, fear, sadness and negative response reviews are very less as compare to anticipation, joy, trust and positive reviews tweets. This clearly showing that customers have liked the service quality, physical attributes and amenities of the Indian five star hotels.

Polarity Analysis (Python)
Sentiment analysis process done by using the Python under the process of analysis data cleaning is done and then data analysis will carried out by using the three different approach.

TextBlob Analysis
Sentiment and polarity analysis in the python can be performed by using libraries and tools. TextBlob library is used for the sentiment analysis
TextBlop is used for sentiment analysis of reviews given by the customers on the TripAdvisor under is approach NLP (Natural Language Processing) and NLTK (Natural Language Toolkit). NLP is a method to make data usable for the system and NLTK is a English based python package. Through TextBlob polarity and subjectivity will calculated as given below:

Polarity
TextBlob output floating range (-1.0, 1.0) 1.0(Positive polarity), -1.0 (Negative polarity) and 0 (neutral evaluation)

Subjectivity
It measures the text whether it a fact or opinion based. Floating range (0.0, 1.0), 0 is a very objective sentence and 1 is very subjective.

Senti WordNet Analysis
SentiWordNet uses the WordNet database. NLTK module includes the English WordNet with 155287 words and 117659 synonym sets that are logically related to each other.It work on POS, lemma of each word and these are used to obtain the synonym sets(synsets). Through this positive, negative, objective scores for all the possible synsets.

VADER Analysis
Vader stands for Valence Aware Dictionary and Entiment Reasoner.It gives result of four dictionary keys: Positive, Negative, Neutral and Compound along with the intensity of emotion.
Sum of (Positive, Negative and Neutral is 1 or near to 1)
Compound Score used to find sentiment
Here is how to read the measurements:
 -1: extremely negative
 1: extremely positive
 0: Neutral or N/A

$$\text{Compound Score} = \frac{X}{\sqrt{(x^2 + \alpha)}}$$

Where x = sum of valence scores of constituent words,
α = Normalization constant (default value is 15).
X calculated by the value of Polarity, Subjectivity and Intensity.
-1 < Compound Score < +1

|	Cleaned Reviews	Lemma	Subjectivity	Polarity	Analysis	Vader Sentiment	Vader Analysis	SWN analysis
1	I wanted a customised meal for myself Staff wa...	want customised meal Staff kind enough call ...	0.671429	0.228571	Neutral	0.8655	Positive	Positive
2	I hosted a office lunch after having couple of...	host office lunch couple good experience hot...	0.620000	-0.172381	Neutral	-0.7645	Negative	Neutral
3	I and my wife gave our wedding reception in We...	wife give wedding reception Welcom ITC Hotel...	0.745000	0.680000	Positive	0.9908	Positive	Positive
4	We held our wedding reception in Welcom hotel ...	hold wed reception Welcom hotel ITC food hig...	0.595833	0.566667	Positive	0.9750	Positive	Positive
5	We attended Welcome Hotels by ITC for my siste...	attend Welcome Hotels ITC sister wed recepti...	0.630952	0.576190	Positive	0.9544	Positive	Positive
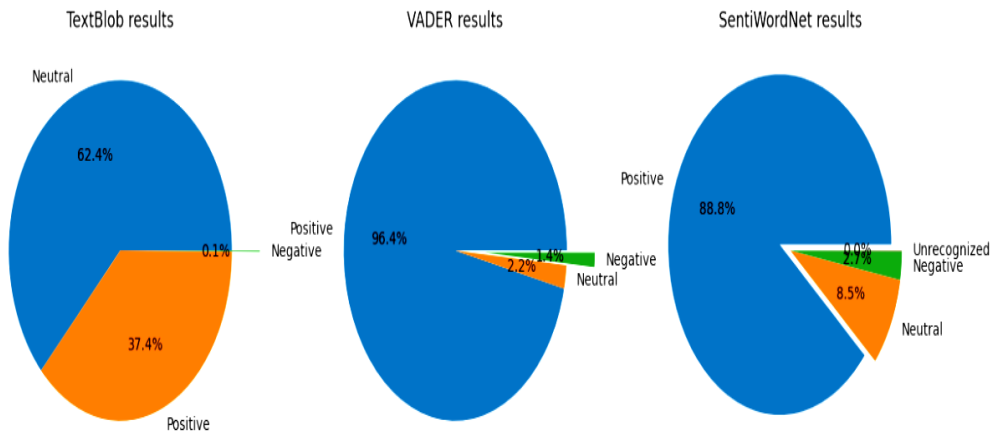
Figure 3. Polarity and sentiment analysis of reviews

The polarity and sentiment analysis through the python on different three parameters as explain above TextBlob results are 62.4% reviews are Neutral, 37.4% reviews are Positive and 0.1% is negative. Vader results are 96.4% are positive, 1.4% is negative and 2.2 are neutral.

The analysis clearly explaining that customer's reviews have liking attitude towards the Indian hotels services as the negative response represented through all the three analysis methods are less as compare with the positive response as shown Figure.3

Figure 4. Positive and Negative Word cloud of reviews

In the text analysis process word clouds are very important to conclude the real results. In analysis process three word clouds are generated positive, negative and bigram word cloud with help of Python as represent by figure.4

In positive word cloud highlighted words are 'comfortable' 'great' 'good' 'amazing' 'well' 'wonderful' 'best' 'helpful' etc. In negative word cloud highlighted words are 'pathetic' 'crowded' 'disappointing' 'failed' 'wrong' etc. Positive word cloud word have higher density as compare negative word cloud that explain customers gives more positives reviews and liking attitudes towards the Indian five star hotels.

Figure 5. Bigram word cloud

Bigram in text analytics also have very important role to analyse and to find the derived meaning of the word used in the word cloud. Words used in bigram cloud

Staff	Hotel	Food	Hospitality	Service
Staff helpful	Room good	Good food	Amazing experience	Service good
Friendly staff	Room Clean	Food delicious	Great hospitality	Service excellent
Thank Mr.	Stay comfortable	Food great	Must visit	Great service
Special thank	Hotel good		Really nice	
	Well maintained		Like home	
	Recommend hotel			

Customers express their experience through reviews and appreciate on different parameters like staff, hotel, food, hospitality and service etc and maximum words are positive it clearly explains that customers like and having favourable attitude towards Indian hotels and their facilities.

6. Hypothesis testing

To find out the influence of distinct attributes rating on the 'Overall Rating' of the Indian five-star hotel. Sample of 100 five-star hotels are selected for data collection from 'Tripadvisor'. For testing the hypothesis correlation between the overall rating and attributes rating is calculated by using the SPSS.

Correlations						
		Overall	Location	Cleanliness	Service	**Value**
Overall		1	.396**	.647**	.740**	.756**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	100	100	100	100	100
**. Correlation is significant at the 0.01 level (2-tailed).						

Hypothesis	Pearson Correlation	Accepted/Rejected
H1: Overall Hotel rating highly influenced by location	.396	Accepted
H2: Overall Hotel rating highly influenced by Cleanliness rating.	.647	Accepted
H3: Overall Hotel rating highly influenced by Service rating.	.740	Accepted
H4: Overall Hotel rating highly influenced by Value rating.	**.756**	Accepted

Value rating is highly correlated with Overall rating as compare to the other attributes rating. It means rating of value have highest influence on the overall rating of the Indian five-star hotels.

7. RESULTS AND DISCUSSION

The study analyzed ratings and reviews of five-star Indian hotels across the major eight cities of India through sentiment analysis. Data was collected from renowned online platform 'TripAdvisor' in the form of hotel reviews. A total no. of 3326 reviews were scrapped and focused on various aspects like rooms, service quality, location and cleanliness etc. The sentiment analysis revealed positive sentiment towards the hotels, indicating that customers are satisfied with service quality, amenities and physical attributes.

The finding describes the significance of online reviews in modifying consumer's perception and influence on the hotel selection decision. The positive sentiment in the reviews suggests high level of satisfaction of customers with the service and facility provided by the Indian five-star hotels. The positive sentiment can contribute towards the brand value, customer delight, customer loyalty and ultimately, establishment success.

The study also examined the impact of various attributes rating on the overall ratings of the hotel rating. Correlation analysis revealed that while different attributes like service, location and cleanliness are the important factors contributing to overall hotel rating, value rating emerged as the most influential rating on the overall rating of hotel. This suggests that customers prioritize value for money when evaluating the hotels.

Study finding have implications for both academic literature and hotel managers practices. Academically, the study contributes to understanding the behaviour and attitude of the consumers and sentiment analysis of the Indian hospitality industry, particularly relate to online reviews. Practically, the insights gained from this research can guide the hotel professionals in formulating strategy and policies to enhance customer satisfaction, improve service quality and optimized pricing strategy to maximize overall rating and enhance the occupancy rate.

8. CONCLUSION

This study investigated the sentiment of the online reviews of the Indian five-star hotels expressed by the customer's across the major eight cities of India, focused on the rooms, cleanliness, service quality and customer value. The findings suggest a predominately positive and favourable sentiment among customers indicating customers are satisfied with the amenities and facilities delivered by the Indian hotels. Moreover, the finding revealed that valued rating have significant influence over the overall rating of the hotel. The findings are valuable for the academicians and hotel management practise for developing the suitable policies and strategies to attract more and more customers as well as also helps to understand the behaviour and attitude of customers. This study emphasizing on shaping the perception of customers in guiding the hotel selection decisions.

9. Limitations

Despite the contribution, several limitations should be acknowledged of this study. India is the one of the tourist destination and fastest growing economy in the world. In this study the analysis relied solely on the data from tripadvisor which may not represent the sentiment of all the customers, the study only considered the five-star hotel customers and neglecting the other customers staying in lower-rated accommodations. Furthermore the tool used in this study may have limitation in accurately capturing the nuances of the customer sentiment. Future research should be the data collected from multiple sources and exploring across the different category of the hotels.

REFERENCES

[1] Akhtar, N., Zubair, N., Kumar, A., & Ahmad, T. (2017). ScienceDirect ScienceDirect Aspect based Sentiment Oriented Summarization of Hotel Reviews. Procedia Computer Science, 115, 563–571. https://doi.org/10.1016/j.procs.2017.09.115

[2] Dina, N. Z. (2020). Tourist sentiment analysis on TripAdvisor using text mining : A case study using hotels in Ubud , Bali, 9(2), 1–10.

[3] Dolnicar, S. (2003). Which Hotel attributes Matter ? A review of previous and a framework for future research, 176–188.

[4] Elango, V., & Narayanan, G. (2002). Sentiment Analysis for Hotel Reviews.

[5] Kanade, P. P. (2019). Automatic Sentiment Analysis of User Reviews, 9(1), 5–10.

[6] Markopoulos, G., Mikros, G., & Iliadi, A. (n.d.). Sentiment Analysis of Hotel Reviews in Greek : A Comparison of Unigram Features, 373–383. https://doi.org/10.1007/978-3-319-15859-4

[7] Pratama, M. D. (2022). Sentiment Analysis User Regarding Hotel Reviews by Aspect Based Using

Latent Dirichlet Allocation , Semantic Similarity , and Support Vector Machine Method, 15(3), 514–524. https://doi.org/10.22266/ijies2022.0630.43

[8]   Priyantina, R. A. (2019). Sentiment Analysis of Hotel Reviews Using Latent Dirichlet Allocation , Semantic Similarity and LSTM, 12(4). https://doi.org/10.22266/ijies2019.0831.14

[9]   Raul, D., & Silva, M. (2019). Sentiment analysis in online customer reviews : The Feels Like Home case Duarte Rodrigues dos Santos Farinas de Almeida, (October).

[10] Rybakov, V., & Malafeev, A. (n.d.). Aspect-Based Sentiment Analysis of Russian Hotel Reviews.

[11] Sanwal, H., & Kukreja, S. (2019). Design Approach for Opinion Mining in Hotel Review using SVM With Particle Swarm Optimization (PSO), 8(09), 519–522.

[12] Sharad, D. A., Ashok, D. S., Dattatray, D. P., & Bhagwat, M. P. (2018). Sentiment Analysis of Hotel Review, 7(11), 174–176. https://doi.org/10.17148/IJARCCE.2018.71139

[13] Stuart, K. R., Mednick, M., & Bockman, J. (n.d.). Structural Equation Model of Customer Satisfaction for the New York City Subway System, (00), 133–137.

[14] Sufi, T. and Singh, S., 2016. Online and offline hotel ratings: a case study of Indian hotels. International Journal of Applied Business and Economic Research (IJABER), 14(4), pp.2555-2574.

[15] Gunasekar, S. and Sudhakar, S., 2019. Does hotel attributes impact customer satisfaction: A sentiment analysis of online reviews. Journal of Global Scholars of Marketing Science, 29(2), pp.180-195.