Hotel Room Allocation Model for Maximizing Positive Reviews: A Sentiment Analysis Framework

Charlie Martin

Word Count - 7760

Contents

Abstract
The objective of this project is to provide new and meaningful insights into the literature surrounding sentiment analysis in the travel and tourism industry and hotel room allocation strategies. This project aims to achieve this by suggesting a new perspective for analysing reviews, specifically hotel reviews, where reviewers are separated by relevant booking features and analysed individually. By threading together sentiments expressed within reviews, the intricate art of room allocation, and the ever-evolving hospitality landscape, this project endeavours to contribute a more comprehensive understanding of these domains. Notably, the outcomes are projected to offer invaluable advantages to front-of-house personnel by infusing them with real-time insights that facilitate informed decision-making within the immediate operational context. This is a stark contrast to the previously exclusively focusing on long-term strategies, but when used in conjunction with such strategies, this approach aims to empowers day-to-day resource allocation with newfound depth.

To achieve these goals, the project takes a quantitative approach, using various statistical measures, programming techniques and machine learning algorithms to reach certain results. The study will conclude that booking type, customer type, duration and room type are all significant factors relating to review sentiment. These are then separated, and the review content analysed separately, uncovering trends and insights that were previously hidden when analysing reviews as a whole. Furthermore, the study concludes that in the dataset, reviews are weighted with a total of 14 scores, 0-13, with 0 being the most likely to leave a positive review, and 13 being the least likely. The most occurring score in the dataset was a strong score of 3, further supporting the strength of the European tourism market, and highlighting the need for studies like these that take the analytics one step further, to gain a competitive edge in an increasingly competitive market.

Introduction
Online reviews are of increasing importance for the hospitality and tourism industry. This is supported in the work of Schuckert, Liu and Law (2015), who found that online reviews play a critical role in sales for the industry. To add to this, Kaplan and Haenlein (2010) concluded that online reviews have become the word of mouth of the digital age. In the words of Duan et al. (2008), word of mouth has long been recognised as the most influential resource of transmission in society. Clearly, the importance of online reviews as a resource has been well established. For hotels, extracting maximum information and value from this resource can provide a competitive edge, but extracting meaningful insights from a catalogue of unorganised personal experiences is no mean feat. This project aims to create a framework for extracting maximum insights from online reviews and develop previous studies further by identifying significant factors in customer profiles relating to sentiment. Using these insights, the project will suggest a room allocation model that weighs bookings based on the likelihood of a positive review compared to a negative one. Several managerial applications and implications from the results will be discussed, with the aim of maximising positive reviews to in turn maximise hotel resource management and revenue.

The main issue this project aims to tackle is the one-dimensional approach to hotel reviewer analytics, which fails to recognise that the vast differences in hotel guests, their expectations, objectives and review process could significantly impact results.

This project follows a quantitative, positivist approach and applies a deductive reasoning pathway. The main analysis of this project is conducted with statistical tests and machine learning algorithms, unsuitable for a qualitative approach. While much of the raw data is in the form of qualitative reviews, the analytical methods convert these to quantitative measures for more reliable results.

This introduction will be followed by a literature review, covering all the relevant literature surrounding the project, and which gaps have been identified for the basis of this project. Following the literature review is the methodology section, which contains information about the data, its justification for use, and any pre-processing changes to the dataset, followed by the methodology for the analysis. After methodology, the findings section clearly outlines the results of the analysis and the accompanying findings, followed by the discussion section which contains the managerial implications and recommendations for future work. The final section is the conclusion, which will summarise the project and its findings. For supplementary material referenced in the text, the appendices are after the conclusion.

Literature review

Review Impact

To start the literature review, this first section will go into further detail of the importance of online reviews. The importance of reviews has been discussed and analysed in many industries. For example, in a study of the restaurant industry by Zhang et al. (2010), where it is noted that online reviews are now an emerging phenomenon and are playing an increasingly important role in consumer purchase decisions. This study concluded that consumer generated ratings are positively associated with the popularity of said restaurants, increasing their reputation and revenue, and had a focus on a certain reviewer attribute, being a consumer, or a critic. These attributes were found to significantly affect the impact of the review. To take this further, Lee and Ro (2016) studied the ability of online reviews to change consumer attitudes and concluded that online reviews are so powerful they can change consumer attitudes from previous experiences. This study also took consumer attributes into account, namely their prior experience and knowledge levels.

In Zhu and Zhang's (2010) work, the goal of the study was to provide positive reconciliation of previous mixed results. This study was on the video game industry specifically, and concluded again that both product and consumer characteristics moderate the influence of online consumer reviews on product sales and suggest that previous contradictory results are likely due to data deficiencies and variations. This conclusion is widely accepted, being cited over 3000 times since.

The impact of online reviews in the hotel industry cannot be overstated. As found by Vermeulen and Seegers (2009) 84% of hotel bookers had their choices impacted by what they saw in online reviews. This study also noted that all reviews, both positive and negative, benefited the hotel via increased brand exposure. This is contradicted by the work of Sparks and Browning (2011), in which it is found that consumers tend to be more influenced by early negative information, particularly if it is easy to process. However, there is more evidence to support that more reviews are always better, such as in the study on Belgium hotels by De Pelsmacker (2018) where it is found that review volume drives room occupancy.

The literature here is providing the consensus that reviews are of increasing volume, impact, and importance for businesses. It is also evident that there is a trend of the inclusion of reviewer attributes within review analysis. From restaurants to gaming, many sectors have included consumer attributes in review analysis, with significant results.

Fake Reviews

For businesses, user reviews can play a significant role in determining the revenue for an organisation, which has in turn increased the incentive for malicious behaviour regarding reviews.

This is confirmed in the work of Mohawesh et al. (2021), whose study on reviews found they directly impact a company's reputation and profitability. This study also notes the dangers of fake reviews often used to exploit consumer purchasing decisions. The prevalence of fake reviews is supported in another study by He et al. (2022), in which an extensive study is carried out into the market for fake reviews on amazon, which is found to be growing. This study also finds that these fake reviews have a significant impact on short term product and company performance. The literature here also highlights that there are two main types of fake review strategies, review bombing and review boosting. Review bombing consists of targeting competitors and spamming their reviews with fake, negative reviews. the goal of this is to reduce the company's performance and lower their reputation with the aim of taking their market share. Review boosting on the other hand is flooding a company or product review with overhighly positive reviews. The goal of this is to raise the overall score, and in turn the reputation of company and increase performance via increased sales. With the growing problem of fake reviews impact company performance in addition to analysis results, various fake review detection methods have been studied. Elmogy et al. (2021) proposes a supervised machine learning method using a combination of natural language processing and the K nearest neighbour classification model to detect fake reviews. This study finds the approach to be successful in fake review detection. This strategy is also shown to be successful in the work of Salminen et al. (2022), who successfully deploy classification techniques to detect fake reviews with near perfect accuracy.

The aforementioned studies, however, detected fake reviews successfully by having pre-identified fake reviews used to train a machine learning model, which was then deployed onto the test set. In the dataset for this analysis, there are no identified fake reviews, or 'verified' information regarding the data, thus fake reviews must be manually identified via certain characteristics before any machine learning models can be deployed. This conclusion is also reached in the work of Dematis et al. (2018), who note that the lack of a globally reliable, annotated training set has shifted the research regarding fake reviews to focus more on identifying indicators and characteristics. In this study, several indicators were discussed regarding spam or fake reviews, including text similarity/duplicates, number of reviews per reviewer, review length and time between reviews. Using these indicators, the study was successful in training a model to identify over 2000 fake reviews. The work of Fontanarava et al. (2017) goes further into the textual analysis regarding fake reviews, successfully using features such as ratio of capital letters and ratio of exclamation sentences (ending with !) to identify fake reviews. These characteristics indicating fake reviews is also supported in the work of Wang et al. (2022), whose study also emphasis the importance of analysing reviewer behaviour in addition to review content, such as leaving many reviews in a short time frame. The literature here regarding fake reviews emphasises two key points. Firstly, fake reviews are extremely likely to be present in the dataset, and thus must be removed for the integrity of the analysis. Secondly, machine learning models are highly effective at identifying fake reviews even when no annotated training set or verified information is available, highlighting key characteristics to identify within the dataset to successfully remove fake reviews.

Sentiment Analysis

Sentiment analysis is a relatively new term and is one of the fastest growing research areas in computer science, partly due to online reviews. As noted by As Mäntylä, Graziotin, and Kuutila (2018), the availability of subjective texts online has caused an outbreak in computer-based sentiment analysis, with 99% of related papers being published after 2004. This study provided a review of some of the most cited and influential studies in the field, some of which will be discussed. One of the most cited papers on sentiment analysis is Mining and Summarizing Customer reviews. In this study, Hu and Liu (2004) mention some of the key problems with the increasing popularity of reviews, namely too many reviews to analyse and reviews being too long to digest. This study discusses techniques for counteracting these problems, such as sentiment classification and feature selection with machine learning. Hu and Liu conclude that the proposed techniques are very promising and will be of increasing importance in the future for both consumers and manufacturers.

A key feature of sentiment analysis is to differentiate between the review sentiment and review rating. This is found in the work of Pang et al. (2002) which found that a crucial characteristic of the rapidly growing online discussion sites is the overall sentiment. This is further supported by Hu et al. (2014) where it is suggested that the previous focus on purely numeric ratings is due to the difficulty of extracting sentiment. In this study of amazon reviews, it is found that the review ratings in fact have no direct impact on sales, but indirectly impact sales through the sentiment. These studies highlight the need for sophisticated, computer-based analysis on sentiment to utilize the growing resource of reviews. This study also found that sentiment classification can be beneficial specifically in business intelligence applications, with machine learning techniques out-performing human baselines. One critical part of applying this kind of analysis is feature selection, which will be discussed next.

Feature Selection and Machine Learning Algorithms

Feature selection is a crucial part of sentiment analysis, as found in many studies. For example, Harish and Revanasiddappa (2017) stated that feature selection is a strategy that can be used to increase categorization accuracy, effectiveness, and computational efficiency. Furthermore, in a review of the literature, Asgher et al. (2014) found that feature proper feature selection techniques play a key role in identifying relevant attributes for analysis and increasing accuracy. This review also found that feature selection is usually categorized within 4 main types, with NLP (natural language processing) being the most common. This is the only category of feature selection relevant to this literature review, and thus the only one mentioned.

One study by Jing et al. (2002) notes the effectiveness of TF-IDF for feature extraction, with an accuracy of 88% using Naïve Bayes, compared to 76% using IDF. Bag of words is one of the most common feature extraction methods. As noted by Qader et al. (2019) its strength lies in its simplicity and therefore applicability to a wide range of classification models while still producing strong results. Building on this, an extensive study was carried out by Ahuja et al. (2019) where feature extraction methods were tested with various machine learning algorithms and concluded that TF-IDF performs on average 3-4% better. This study also concluded that logistic regression performed better for sentiment analysis, which is supported in a similar study by Prabhat and Khullar (2017). With respect to the literature discussed, a TF-IDF feature selection approach will be adopted, and tested on various machine learning algorithms.

Sentiment Analysis in the Travel and Tourism Industry

Sentiment analysis is certainly a popular topic within this industry, and many studies have been carried out. Generally, the focus of these studies is to identify key features of both positive and negative reviews and prescribe advice to hotels accordingly. Results vary between studies, for example Tran et al. (2019) found that location and restaurants were the two most important aspects for positive reviews, which is contradicted by Sodanil (2016), where it was found that the room was significantly more impactful on positive reviews than restaurants. Most notable for this project is a study by Chang et al (2020) on luxury hotel reviews. As mentioned in this study, customers for luxury hotels tend to focus more on details as upscale amenities are expected, making these reviews the ideal dataset for such analysis. It is also mentioned in this study that future research may benefit from further investigation of customer profiles, which is a gap in the literature of this field in general, as previously noted. Hence, one of the goals of this project.

Hotel Room Allocation Models

Hotels are in a unique position of seeing customer profiles before they arrive, allowing for additional opportunity to manage resources, namely room allocations. This had led to the development of various models to optimise this. The literature surrounding this notes the importance of room allocation, and even refers to it as a "statistics game" in a study by Song et al. (2010). This study proposes a static game model to the solve the hotel room inventory problem, and notes the important managerial implications on revenue management from optimal allocation. Additionally, Aydin and Birbil (2018) propose a mathematical model to optimize hotel room allocation with long stay booking. This study focused on maximising revenue by finding the optimal number of rooms reserved for walk ins compared to stay-over guests. In the work of Ni et al. (2020), a model is developed to optimise the different sale channels a hotel may have, such as various booking sites and direct bookings. This model aims to optimise revenue by minimising regret in the booking varied booking prices from different channels, suggesting a model that allows hotels to accept or reject bookings from channels immediately while minimising risk. A different model is proposed in the work of Qin (2014), where there is a focus on consumer choice and feedback driving the model. This model is dictated by individual room attributes, including layout, price, lighting which are used as predictors in analysing the rooms demand. This model successfully calculated the level of demand for each room but was noted as being particularly complex and required further study into time dependant booking trends and market competition.

The literature surrounding hotel room allocation thus far is primarily focused on long term business strategies to optimise hotel revenue and resource management, with models having varying focus on rooms, prices, booking channels and consumer choices. Despite the variance, these studies all note that correct hotel allocation and resource management can benefits hotels via increased profitability and note the importance of prioritizing bookings accordingly. Linking back to the review impact section, one study mentioned found that review volume drives room occupancy, which could significantly impact a room allocation model. Thus far, there have been no links between hotel room allocation and the new resource of online word of mouth. This project will aim to create a quick and practical framework for hotel managers to use, where all bookings are weighted and thus prioritised based on the results from the sentiment analysis. This will allow hotels to allocate rooms accordingly to maximize positive reviews, or minimize negative ones, which has been shown to benefit the organisations sales, reputation, and profitability. This differs from the previously discussed models as this model will not provide a long-term strategy relating to booking levels or price, but rather a short-term day to day strategy to compliment any existing long-term strategies. This model would work best in conjunction with other models, such as a pricing model and a demand model, to add a final layer of optimisation that can be deployed easily by front of house staff on a daily basis.

Summary

To summarize, the literature discussed has shown that the importance of online reviews has been well established across many industries, and trends show that this is only going to increase in the future. The literature has also shown that with the rise of popularity of these reviews, there has also been a rise in malicious activity related to them, namely fake reviews. This activity has shown to be prominent in many industries and must be considered before analysing the content of reviews. The analysis of genuine reviews has provided many benefits for industries, and an extremely effective way to analyse these is via sentiment analysis. The studies discussed have emphasized the capabilities of sentiment analysis, with machine learning techniques and feature selection combinations having high rates of success in these cases. The sentiment analysis research into the tourism industry specifically has yielded promising results with significant managerial implications applying the insights and information gained from the research. However, the literature thus has a trend viewing all hotel consumers as just that, consumers; failing to recognise that different customer types may have different needs even with the same hotel. This gap in the literature is one this project aims to gain an insight into. Furthermore, the literature has established hotel room allocation is a significant aspect of hotel revenue and resource management of which the importance cannot be understated. Much of the work in this area focuses on long term allocation models to maximise revenue, minimise lost revenue or prioritising booking channels. However, there has yet to be a link between the online reviews and the insights within and any form of room allocation model. A sentiment analysis investigation may be useful in gaining further insights from reviews to drive an allocation model and join these two bodies of literature.

Methodology

Dataset Background and Justification

This section will provide a brief introduction to the dataset, how the data was collected and what features it contains, followed by any changes made to the dataset. Firstly, the dataset is a publicly available dataset of hotel reviews found on Kaggle. The data was web scraped from popular hotel booking site booking.com and contains data for over 515,000 reviews on 1493 hotels across Europe. Booking.com allows reviewers to leave both a positive and negative review, so entries contain these alongside a final score, some booking information and hotel information. For the full data schema and information, see **appendices figure 1**.

The dataset is suitable for this analysis as Europe has a relatively large and varied tourism market and plays a leading role in the tourism sector (Bürgisser and Di Carlo, 2023). Being in a leading market naturally creates a competitive environment, and hotels need to keep pace with trends and technological advancements to gain competitive edges, providing sound motivation for sentiment analysis on this dataset. This also adds the potential applicability of any findings to other tourism markets, which may be in a similar state for comparison, or further behind for prediction. Furthermore, booking.com's excellency as a source has been noted in the past, such as in the work of Mellinas etl al. (2015), who note the website's ability to reliably collect hotel reviews in an inexpensive and convenient manner.

Firstly, the dataset is checked for missing values and resulted in 166 null entries in both the lat and lng columns. For this project, these columns will not be used for any analysis and thus the entire columns are dropped. Secondly, the 'Sentiment' column is created based on the reviewer score, for use in future classification models. It contains:

- Positive – Score >= 7
- Neutral – Score >= 5 and <= 6.9
- Negative – Score < 5

The columns 'Tags' is then split into 4 separate columns based on the information within. These include 'Booking type', 'Customer type', 'Room type' and 'Duration'. These columns contain a substantial amount of different data entries with each hotel having some variation in entry names. The data within each column is grouped into several relevant categories. **Appendices figure 2** contains all the categories within these columns and what data went into them. These 4 columns make up the features for customer profiles that will be examined for significance in relation to review sentiment.

Fake reviews

Before the main analysis, the strategy for fake reviews must first be developed and deployed to ensure the impact of fake reviews is mitigated as much as possible. With respect to the literature previously discussed, the following characteristics have been selected as potential fake review indicators.

- Duplicate Review Text
- Similar Review Text
- High Volume of Reviews
- Review Length

Firstly, the dataset is filtered to show duplicated reviews. This results in a high volume of data, mostly containing common phrases for reviews, such as "no negative". To further filter results the dataset is filtered for duplicated reviews with 75 or more characters as this further reduces the likelihood that these reviews were duplicated coincidently. This resulted in a total of 782 reviews, which are added to the 'suspicious reviews' data frame. This first investigation already demonstrates evidence of potential fake reviews, as there are 6 duplicated reviews submitted for 1 hotel from 6 different reviewer locations on different dates.

Next the dataset is filtered for reviews with high similarity. Specifically, positive text and negative text reviews over 300 characters which are 90% similar, followed by over 400 characters with an 80% similarity. This resulted in a further 54 positive and 64 negative reviews being added to the suspicious reviews data frame. Finally, high volume reviews are added, where the reviewer has left over 100 reviews. This adds a further 610 reviews to the suspicious review data frame, bringing the total number to 1510.

Figure one shows the distribution of reviewer score amongst the identified suspicious reviews. Referencing the literature, it can be ascertained that fake reviews will fall into one of two categories – review bombing or review boosting. This means that reviews with a neutral score are likely 'false positives' – genuine reviews that happen to contain one or more of the outlined characteristics, and thus should be removed. Removing these entries left a total of 1371 suspicious reviews.

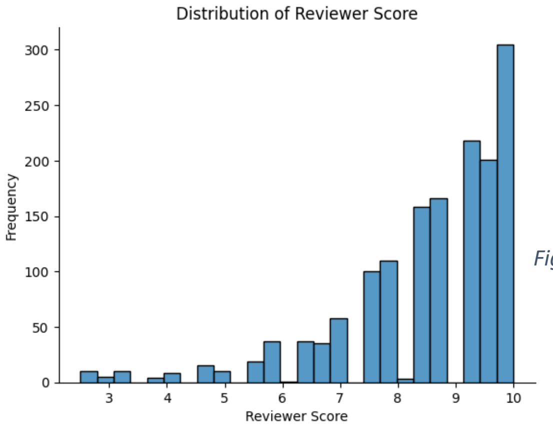

Figure 1 - Distribution of Reviewer score

To determine if the content of the suspicious reviews was comparable to that of the remaining reviews, several machine learning classification models were built. This includes a logistic regression (LR), naïve bayes (NB) and random forest (RF) with a TF-IDF feature selection method. Firstly, the 'Positive_review' and 'Negative_review' columns were combined into one column, 'ReviewText' and the text was cleaned by removing white spaces, links, email addresses, and special characters. Then a random sample of 1200 suspicious ones were taken, alongside a random sample of 1200 reviews from the dataset that were not identified as suspicious. The column 'fake' was created and populated 'yes' or 'no' accordingly. This sample was split into 'training' and 'test' sets, consisting of 80% training and 20% test.

The machine learning pipeline was built using pyspark, and consisted of tokenizing the words in ReviewText, removing stop words, counting the word frequency, calculating the IDF, preparing the target variable 'fake' and finally selecting the model. **Figure 2** shows the results of all the machine learning models on the test sets. For the full parameters selected for each model, see **appendices figure 3**. Due to the small sample size in comparison to the dataset the best performing model (LR) is replicated for a total of 5 times, with 5 different samples from the dataset. The model performed consistently throughout and can successfully differentiate between the suspicious reviews and the data sample using the review text. The ML results in combination with literature supporting the suspicion of the outlined characteristics are strong evidence that these reviews are fake, and thus removed from the dataset.

	LR	NB	RF
Accuracy	0.732	0.651	0.623
F1	0.731	0.648	0.62

Figure 2 - Machine Learning Results

Customer Profiles

As previously discussed, this study aims to distinguish what available customer information is significant to review sentiment to outline possible features to base an allocation model on. In this case, a feature is significant if review sentiment is not independent from said feature. To test for this the chi-squared test will be deployed on each feature outlined. The chi squared test is a method for analysing the association between two variables, and its robustness and effectiveness has been noted in several studies, such as in that of Rana and Singhal (2015) and McHugh (2013). The tests will be carried out on a balanced sample, containing 5000 positive, negative, and neutral reviews. For each chi square test carried out, the test statistic and p value will be analysed to determine statistical significance to a 95% confidence level.

The hypotheses for these tests are as follows:

- H0 – There is no association between booking type and review sentiment.
- H1 – There is an association between booking type and review sentiment.

- H0 – There is no association between customer type and review sentiment.
- H1 – There is an association between customer type and review sentiment.

- H0 – There is no association between room type and review sentiment.
- H1 – There is an association between room type and review sentiment.

- H0 – There is no association between duration and review sentiment.
- H1 – There is an association between duration and review sentiment.

Once the significant customer profile features have been established, the text from each will be further analysed for additional insights. Before the text can be analysed, it should be established that the text within does accurately portray review sentiment. To achieve this, classification models will be deployed with the aim of correctly classifying review sentiment based on the review text. A LR, NB and RF model will be built with the same parameters outlined in **appendices figure 3.** The model is built with a balanced sample taken randomly from the dataset, consisting of 5000 positive and 5000 negative reviews.

Allocation model

The room allocation model will adopt a sophisticated approach to optimizing guest satisfaction by leveraging the likelihood of a review being positive based on the previous entries with similar customer profiles. Each significant customer attribute will be analysed to see which specific entry type is most likely to leave a positive review, creating a multi-faceted scoring system taking into account all relevant attributes. For example, if business bookings were 50% positive, 50% negative, but leisure reviews were 55% positive, 45% negative, therefore leisure reviews will be weighted more favourably than business reviews. The scoring for each category of customer profile will start at 0 (most likely to leave positive) and continue up in 1 for each unique category. The final score for the booking will be a combination of the results from each significant attribute, with a perfect score of 0 meaning that this booking type is the most likely to leave a positive review, and increasing score indicates increasing likelihood in leaving a negative review.

Findings

Fake reviews

This project was successful in identifying and removing fake reviews from the dataset. The results shown in **figure 2** support the conclusion that fake reviews are present in the travel and tourism industry throughout Europe. These results also support the work discussed regarding fake review features, and it can be concluded that review length, review duplication, review similarity and volume of reviews are strong indicators of suspicious review activity. Furthermore, it can also be ascertained that within the European hotel industry review boosting is significantly more common that review bombing. This is evidenced in **figure 3** which highlights the substantial difference in counts for positive and negative reviews in the identified fake reviews, with positive reviews being significantly more prominent.

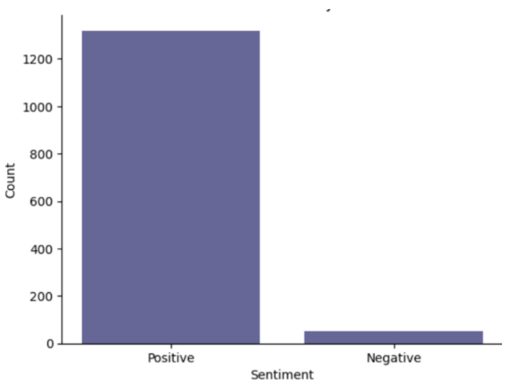

Figure 3 - Fake review sentiment percentages

Customer Profiles

To test for association between review sentiment and booking type, customer type, room type and duration, chi squared tests were carried out. **Figure 4** contains all the test statistics and p values for the individual tests.

Feature	Test statistic	P value	Significant
Booking type	236.6	<0.001	Yes
Customer type	120.5	<0.001	Yes
Room type	82.9	<0.001	Yes
Duration	40.5	<0.001	Yes

Figure 4 - Chi square results

As shown, the test on booking type had a test statistic of 236.6 and has a degree of freedom of 2. Using the chi squared distribution table, it can be seen that 236.6 > 5.99. Furthermore, the p value is <0.001 which is less than α (0.05), therefore the null hypothesis is rejected, and it is concluded that sentiment is not independent of booking type.

The test on customer type had a test statistic of 120.5 and has a degree of freedom of 8. Using the chi squared distribution table, it can be seen that 120.5 > 15.51. Furthermore, the p value is <0.001 which is less than α (0.05), therefore the null hypothesis is rejected, and it is conclude that sentiment is not independent of customer type.

The test on room type had a test statistic of 82.9 and has a degree of freedom of 12. Using the chi squared distribution table, it can be seen that 82.9 > 21.03. Furthermore, the p value is <0.001 which is less than α (0.05), therefore the null hypothesis is rejected, and it is concluded that sentiment is not independent of room type.

The test on duration had a test statistic of 40.5 and has a degree of freedom of 6. Using the chi squared distribution table, it can be seen that 40.5 > 12.69. Furthermore, the p value is <0.001 which is less than α (0.05), therefore the null hypothesis is rejected, and it is concluded that sentiment is not independent of duration.

These results establish that all four of the outlined customer features are significant for review sentiment, and thus should be included in the room allocation model. To gain further insights into each feature, the review text will be analysed. To establish the review texts significance to the review sentiment, several classification models were built. The results are shown in **figure 5,** with the best score being a logistic regression model with an accuracy of 0.832 and a F1 score of 0.8317. From these results we can ascertain that review sentiment can correctly be classified using only the review text, indicating its effective portrayal of customer sentiment. Thus far then, it has been established that booking type, customer type, room type, duration and the review text are all significant factors regarding review sentiment, and the text will now be analysed for further insights.

	LR	NB	RF
Accuracy	0.832	0.822	0.773
F1	0.8317	0.82	0.76

Figure 5 - Machine learning results (sentiment)

Text analytics

Appendices Figure 4 shows the most occurring words in both positive and negative reviews. As shown, the most occurring words for both positive and negative reviews are room, staff, location, and breakfast. Comfortable and clean also appear in the positive reviews, indicating that the rooms are the main driving factor for positive reviews. However, words such as dirty, bathroom and shower occurring a lot in negative reviews indicate that the lower quality rooms drive the negative reviews. It can also be seen that money only occurs in the negative review section, indicating that not only do poor prices promote negative reviews, but good prices are less likely to be mentioned in positive ones.

To gain further insights, bigrams were constructed on all the popular and negative reviews, with the results shown in **appendices figure 5**. From the bigrams, it can be further concluded that the price and value for money are influencing factors for negative reviews, along with customer service, poor breakfast, and rude reception/staff. The positive review bigram indicates that friendly staff and clean and comfortable rooms are indicators for positive reviews. Furthermore, the positive review has a lot of indicators related to the hotel location, including its proximity to transport options.

The review text was also analysed for each booking type to identify and variances or trends. Firstly, leisure bookings (**appendices figure 6)** were compared to business bookings (**appendices figure 7)**. These remained largely similar, with location, staff and comfortable room occurring in the positive reviews. The negative reviews remailed largely similar, with noise and cleanliness appearing in both. However, business bookings also had many negative reviews containing “free Wi-Fi”, indicating that the free Wi-Fi, or lack of, is particularly problematic for business bookings. The reviews from couples and solo travellers were also compared, and similarly contained little discrepancies in positive reviews, with free Wi-Fi being a factor only for solo traveller negative reviews.

On comparison of room types, it was found again that positive reviews in particularly contained little discrepancy throughout, but there was differences in hotel features that different bookings types took note off. For example, in the analysis of the bigrams for luxury bookings (**appendices figure 8),** housekeeping was a common phrase for negative reviews, unlike the others. Additionally, the bigrams for family bookings (**appendices figure 9)** show that swimming pool are mentioned often in both positive and negative reviews, indicating that family bookings value a swimming pool much more than other booking types.

Comparing the different durations highlights that there are differences in experiences and expectations for the various stay lengths. The bigrams for long stay bookings (**appendices figure 10)** show that these guests valued house keeping highly for positive reviews, and had a greater expectation for the hotel restaurants food as it occurs commonly in negative reviews. In comparison, single night stays (**appendices figure 11)** mentioned breakfast more frequently and was more consistent with both average stays (**appendices figure 12)** and extended stays (**appendices figure 13).**

Allocation Model

The results from the allocation model ranked the bookings from a ‘perfect’ score of 0, or most likely to leave a positive review, to a ‘worst’ score of 13, most likely to leave a negative review. The combination of reviewer attributes most likely to leave a positive review is ‘Couple’, ‘Leisure trip’ ‘Average stay’ in a ‘Suite’ and occurred 4314 times in the dataset. The combination most likely to leave a negative review was ‘Solo traveller’, ‘Business trip’, ‘Long stay’ in a ‘Single room’ and occurred 206 times in the dataset. The most occurring booking combination, which was ‘Couple’, ‘leisure trip’, ‘Average stay’, in a ‘Double room’, had a final score of 3, making it the 4th most likely to leave a positive review. Score 0 and score 13 were also the only scores with a singular unique combination, with all other scores have several combinations scoring the same. **Figure 6** shows all the possible combinations that would score a 6.

```
Score 6:
  Customer Type: Solo traveler, Booking Type: Leisure trip, Duration: Average stay, Room Type: Double Room
  Customer Type: Group, Booking Type: Leisure trip, Duration: Average stay, Room Type: Twin Room
  Customer Type: Group, Booking Type: Business trip, Duration: Single night, Room Type: Double Room
  Customer Type: Solo traveler, Booking Type: Business trip, Duration: Single night, Room Type: Luxury Room
  Customer Type: Couple, Booking Type: Leisure trip, Duration: Single night, Room Type: Twin Room
  Customer Type: Group, Booking Type: Leisure trip, Duration: Extended stay, Room Type: Double Room
  Customer Type: Family, Booking Type: Leisure trip, Duration: Average stay, Room Type: Family Room
  Customer Type: Group, Booking Type: Leisure trip, Duration: Single night, Room Type: Family Room
  Customer Type: Family, Booking Type: Leisure trip, Duration: Single night, Room Type: Double Room
  Customer Type: Couple, Booking Type: Business trip, Duration: Average stay, Room Type: Twin Room
  Customer Type: Couple, Booking Type: Leisure trip, Duration: Extended stay, Room Type: Family Room
  Customer Type: Family, Booking Type: Business trip, Duration: Extended stay, Room Type: Luxury Room
  Customer Type: Solo traveler, Booking Type: Leisure trip, Duration: Extended stay, Room Type: Luxury Room
  Customer Type: Family, Booking Type: Business trip, Duration: Average stay, Room Type: Double Room
  Customer Type: Couple, Booking Type: Leisure trip, Duration: Average stay, Room Type: Single Room
  Customer Type: Couple, Booking Type: Business trip, Duration: Single night, Room Type: Family Room
  Customer Type: Couple, Booking Type: Leisure trip, Duration: Long stay, Room Type: Double Room
  Customer Type: Solo traveler, Booking Type: Leisure trip, Duration: Single night, Room Type: Other
  Customer Type: Solo traveler, Booking Type: Business trip, Duration: Average stay, Room Type: Other
  Customer Type: Couple, Booking Type: Business trip, Duration: Extended stay, Room Type: Double Room
  Customer Type: Family, Booking Type: Leisure trip, Duration: Extended stay, Room Type: Other
  Customer Type: Family, Booking Type: Business trip, Duration: Single night, Room Type: Other
  Customer Type: Group, Booking Type: Business trip, Duration: Average stay, Room Type: Family Room
  Customer Type: Solo traveler, Booking Type: Business trip, Duration: Extended stay, Room Type: Suite
  Customer Type: Group, Booking Type: Leisure trip, Duration: Long stay, Room Type: Other
  Customer Type: Group, Booking Type: Business trip, Duration: Long stay, Room Type: Luxury Room
  Customer Type: Family, Booking Type: Leisure trip, Duration: Long stay, Room Type: Luxury Room
  Customer Type: Solo traveler, Booking Type: Leisure trip, Duration: Long stay, Room Type: Suite
  Customer Type: Group, Booking Type: Business trip, Duration: Extended stay, Room Type: Other
```

Figure 6 - Score 6 combinations

Discussion

Managerial implications

Firstly, there are several potential implications and strategies that can be developed from the results of the model. The two main strategies would be minimising negative reviews vs maximising positive ones. Naturally, every hotel has rooms that vary in quality, despite being the same room type and price. This can be down to different layouts, sizes, or views. When allocating bookings to specific rooms, hotels could opt to strategize into maximizing positive reviews, which would mean prioritising the bookings with a lower score to be allocated first, ensuring these bookings get the optimal rooms. In theory, this then results in an increased number of positive reviews via increased positive experiences for the customer. Alternatively, hotels could opt to minimize negative reviews in this scenario and allocate bookings with higher scores first. This would potentially mitigate the increased likelihood of negative reviews by giving the 'best' room to the least likely to leave a negative review, and the 'worst' room to those most likely to leave a negative review.

This model could also guide strategies depending on the varying occupancy rates of a hotel. During a quiet period when hotels are not at max capacity, it is common practice to upgrade rooms on arrival for guests. Since the room is not booked, and thus becomes revenue that can't be reclaimed, upgrading for the customer comes at no cost to the hotel. However, upgrading the correct customers can still provide further value for the hotel, optimising the revenue management. With the proposed model, rooms could be upgraded to minimise the total score across the bookings, via selecting the highest score rooms to upgrade. Reducing the total score for the hotel maximizes the likelihood of good reviews for the capacity.

This model also has potential to provide advantages in overbooking situations. Overbooking is when a hotel sells more rooms than are available, banking on some bookings being cancelled or not showing up. This is an effective strategy hotels deploy to mitigate lost revenue on late cancellations and is noted in the work of Zhai et al. (2023) as being by far the most common. Occasionally, this leads to scenarios where all bookings show up, and some kind of compensation and out booking is required, which has considerable impact on customer experience and hotel reputation. Various models exist for minimising negative impacts from this exact situation, such as the one discussed in Zhai's work, and include various factors from booking type to time of year. To aid such models, the model could be deployed as an extra layer of filtration when identifying potential bookings to be out booked. Minimising the likelihood of a negative review from this situation will benefit hotels reputation.

The insights from the textual analytics can also be applied directly from hotel managers. The findings have shown that business bookings are more likely to leave a negative review than leisure bookings, and that these bookings have a particular interest in free Wi-Fi and value for money. To remedy this, hotels could offer a business booking package, with a Wi-Fi deal for premium internet. The Wi-Fi was also shown to be particularly problematic across several categories, including single person booking. Business and single person bookings had some of lowest rates for positive reviews, so addressing these issues directly has potential to increase positive review rates significantly.

The results have also shown that a large majority of topics in both positive and negative reviews was similar, despite looking at each significant booking characteristic separately. This indicates that hotels can focus on generalised strategies for standard amenities, like the rooms, breakfast, and room service and continue to receive mostly positive reviews. However, as mentioned previously, the European tourism industry is one of the word leading markets and is naturally very competitive. To gain a competitive edge in such a market could be as small as 5% more positive reviews, and the tailored insights mentioned in this study may be niche but could also provide the small boost hotels need to distinguish themselves. Furthermore, to remain competitive in such a market a hotel must manage trends, and plan. The more specific textual analysis on certain groups will detect trends much quicker and would allow hotels to adjust and strategize accordingly with ample time.

Improvements for further work

Although this study made valuable contributions, there was certainly some limitations to the project, and future work can benefit from the following. Regarding the fake reviews, the future work could be substantially improved by developing a model for detecting fake reviews as outlined in this study, and then deploying it on the entire dataset. This study lacked the computational power to deploy a model on half a million entries. Furthermore, this would require a larger training set and manually identified real reviews alongside the manually identified fake reviews. This would provide better insights into not only characteristics and prevalence of fake reviews, but also refine the results from the sentiment analysis.

This research can be further developed by having a more substantial dataset relating to a specific hotel. This would allow the model to incorporate various other factors and weight bookings accordingly for more accurate results. This includes seasonal factors impacting customer expectations, occupancy rates or price rates. A more in-depth dataset for customer history would also be advantageous, while this dataset did specify how reviews each reviewer had left, there was no way to determine how often customers frequent a hotel. Seeing a specific customers review history would significantly improve the model proposed. Furthermore, having a tailored dataset detailing each of the hotel's rooms and customer feedback on specific rooms, like the one discussed previously in the work of Qin (2014).

Future research would also benefit from a depth analysis into the benefits of maximising positive reviews, compared to minimizing negative ones. If these strategies were tested at various hotels and results collected, it could provide a solid basis for developing further strategies regarding targeting review outcomes. A hotel deploying a strategy for each for an extended period of time and collecting adequate data would create a fine resource for refining models such as the one proposed, alongside developing more long-term business strategies.

Naturally, future research would also benefit from more detailed customer profiles. However, for these to be collected the ethical issues surrounding it must be considered. Personal information requires permission to collect, store and analyse, and ethical approval would be needed for any such activity.

Finally, this model was of course never taken from the experimental phase to the test phase. Deploying the model into a real hotel and measuring results would be extremely beneficial for this study and similar ones.

Conclusion

To conclude, this project successfully identified a substantial number of fake reviews within the dataset. These reviews were manually identified, and the study makes contributions determining review characteristics that could potentially indicate fake or spam reviews. These include review length, duplicated reviews, review similarity and number of reviews from reviewer. Additionally, this study found that a logistic regression classification model performed better than a naïve bayes and random forest with a TF-IDF feature selection method. This was concluded in the classification models for both the fake review detection and sentiment analysis, contributing to the conflicting literature in this subject. The sentiment analysis of this study was successful in classifying reviews into sentiment categories based on review text. The text was analysed and outlined key indicators for both positive and negative reviews, including room, staff, location, and cleanliness.

Furthermore, this study successfully identified 4 reviewer booking features that were significant to review sentiment. That is, review sentiment is not independent of these features, thus this should not be overlooked when performing review analysis. These were booking type, room type, customer type and duration, and were identified with a chi squared test, achieving both significant test statistics and p values. The review text was then analysed for trends and patterns pertaining to each specific category. This concluded that the results were largely similar within categories, but there are some underlying trends, such as families having an emphasis on swimming pools, and business bookings having an emphasis on in Wi-Fi. This contributes to the literature on this topic by further confirming the important factors for positive and negative reviews, but also demonstrating that viewing the reviewers as different groups as opposed to just a reviewer can reveal trends previously hidden, which can be beneficial for hotel resource and revenue management.

The study also produced a basic allocation model for maximising positive reviews. The bookings were weighted based on 4 significant features outlined, dependant on what % of each group leaves a positive, negative, or neutral review. The goal of this model is not to be the defining piece of business strategy, but rather a finishing touch. The model can be used in the short term, day to day business operations by front of house staff, as opposed to the traditionally long-term business strategies models surrounding room allocation. This model, used in conjunction with a sound business strategy, can help provide hotels with further optimisation of resources to gain a competitive edge in an ever-growing industry by empowering staff with real time, actionable insights.

Appendices

Column Name	Description
Hotel Address	Address of hotel
Review Date	Date of review
Average Score	Average score for hotel
Hotel Name	Name of hotel
Reviewer Nationality	Nationality of reviewer
Negative Review	Main text from negative review
Review Negative Word Count	Total word count in negative review
Positive Review	Main text from positive review
Positive Review Word Count	Total word count in negative review
Total Number of Reviews Reviewer Has Given	Number of reviews from specific reviewer
Total Number of Reviews	Number of reviews for hotel
Tags	Additional information from booking, including buisness or leisure and room type
Days Since Review	Days between review and web scrape
Additional Number of Scoring	Score left by guests that opted to not leave a full review
Lat	Latitude of hotel
Lng	Longitude of hotel

Figure 1 - Data Schema

Column Name	Category	Contains
Booking type	Business	Business Trip
Booking type	Leisure	Leisure Trip
Customer type	Solo	All variations of solo and single bookings
Customer type	Couple	All variations of couple bookings
Customer type	Family	All variations of family bookings
Customer type	Group	All variations of group bookings
Room type	Double	All variations of double room types
Room type	Luxury	All variations of deluxe, premiere, luxury room types
Room type	Single	All variations of single room types
Room type	Suite	All variations of suites
Room type	Family	All variations of family room types
Room type	Other	Room types not applicable to the previous 5
Duration	Single night	Single night bookings
Duration	Average stay	Bookings from 2-7 nights
Duration	Extended stay	Bookings from 8-14 nights
Duration	Long stay	Bookings over 14 nights

Figure 2 - Category Schema

Model	Parameter	Variation
Logistic Regression	count_vec.vocabSize	[500, 2000, 4000]
Logistic Regression	idf_cal.minDocFreq	[5, 10, 15]
Logistic Regression	lr.regParam	[0.1, 1 , 2.5, 5]
Naïve Bayes	count_vec.vocabSize	[500, 2000, 4000]
Naïve Bayes	idf_cal.minDocFreq	[5, 10, 15]
Naïve Bayes	nb.smoothing	[0, 1, 2]
Random Forest	count_vec.vocabSize	[500, 2000, 4000]
Random Forest	idf_cal.minDocFreq	[5, 10, 15]
Random Forest	rf.numTrees	[5, 10, 15, 20]
Random Forest	rf.maxDepth	[2, 3 , 4 , 5]

Figure 3 - Machine learning parameters

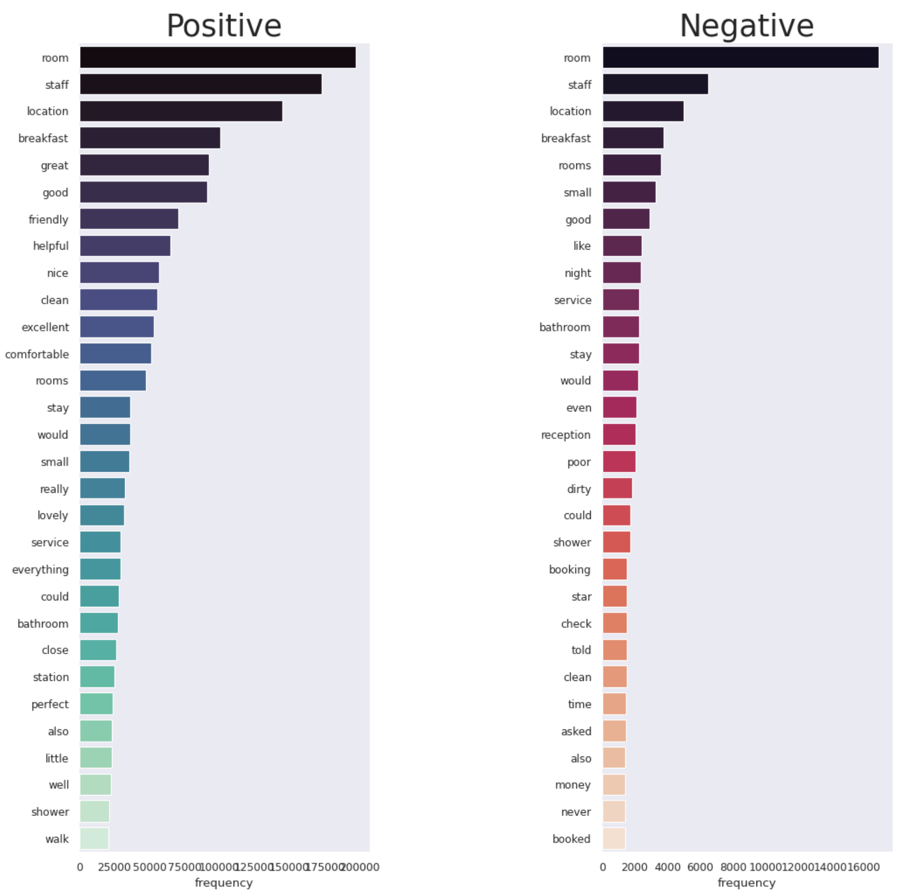

Figure 4 - Most frequent words

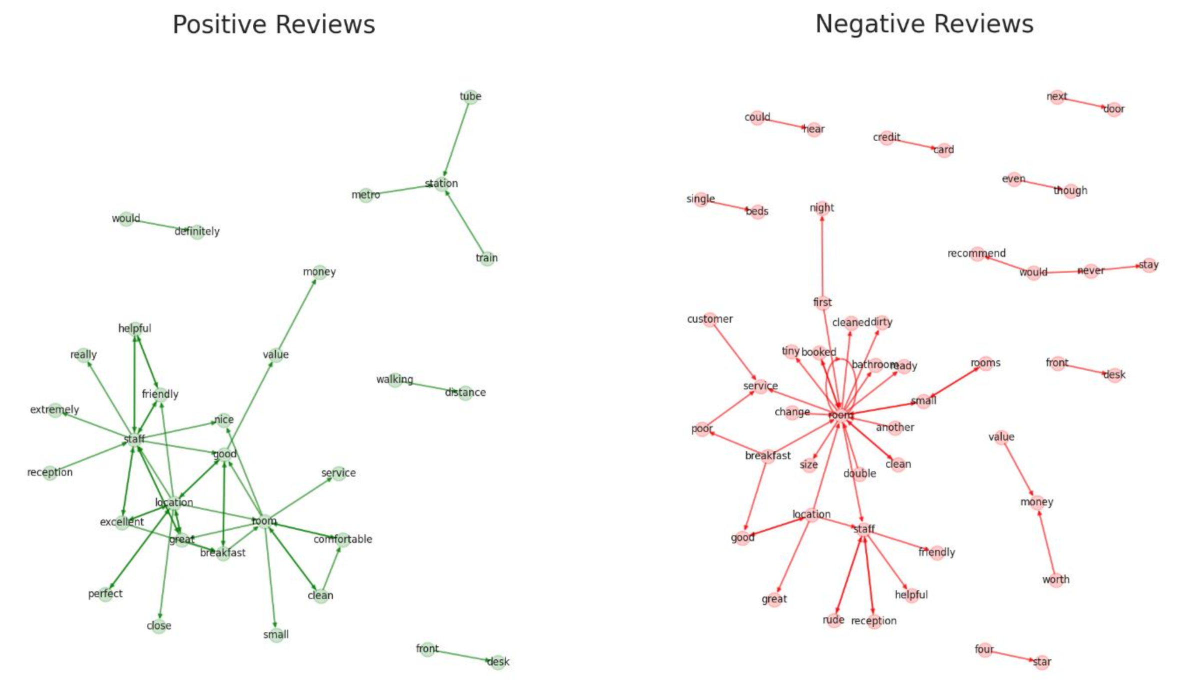

Figure 5 - Bigrams

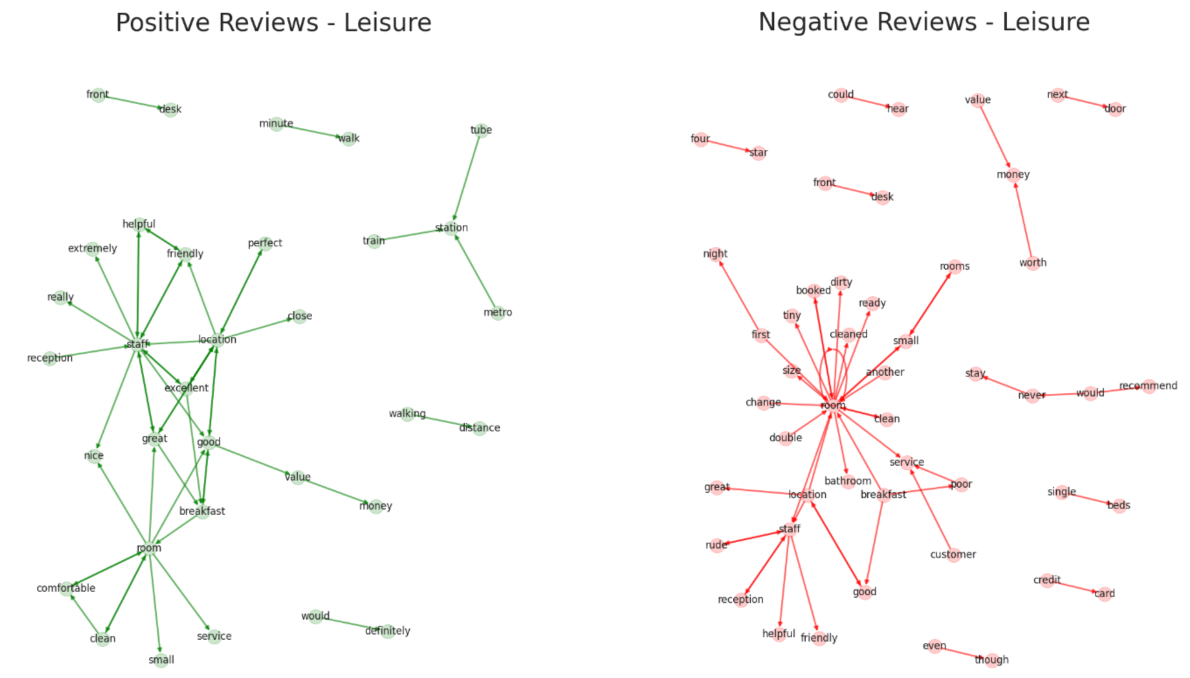

Figure 6 - Bigrams (Leisure)

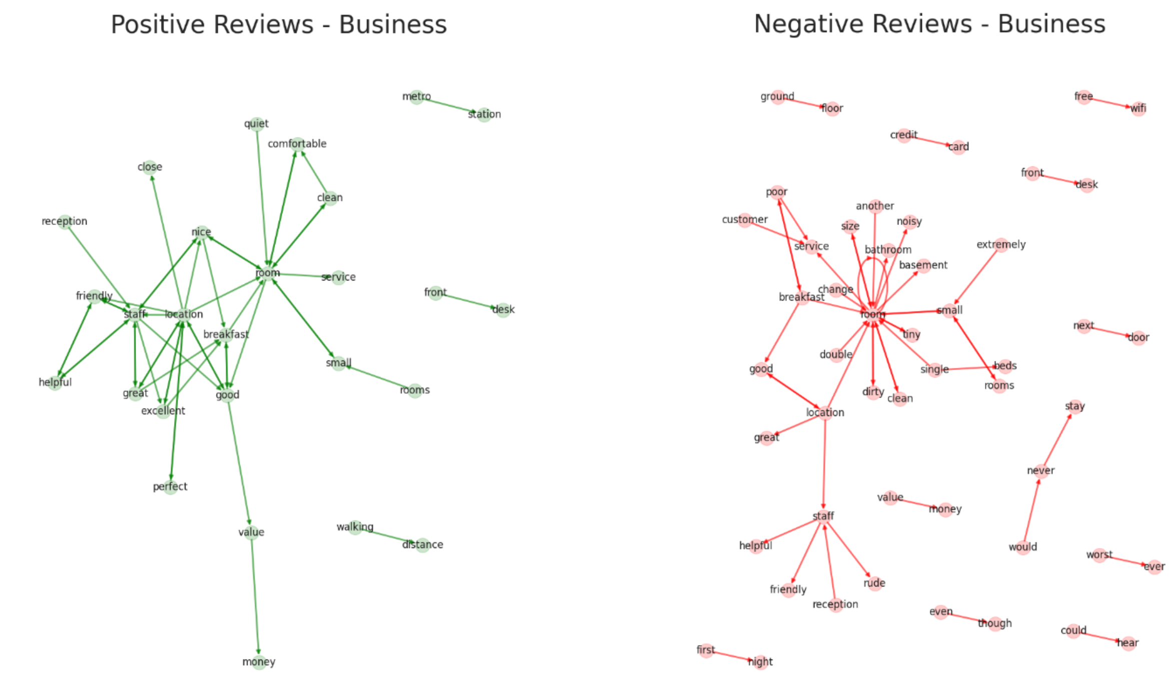

Figure 7 - Bigrams (Business)

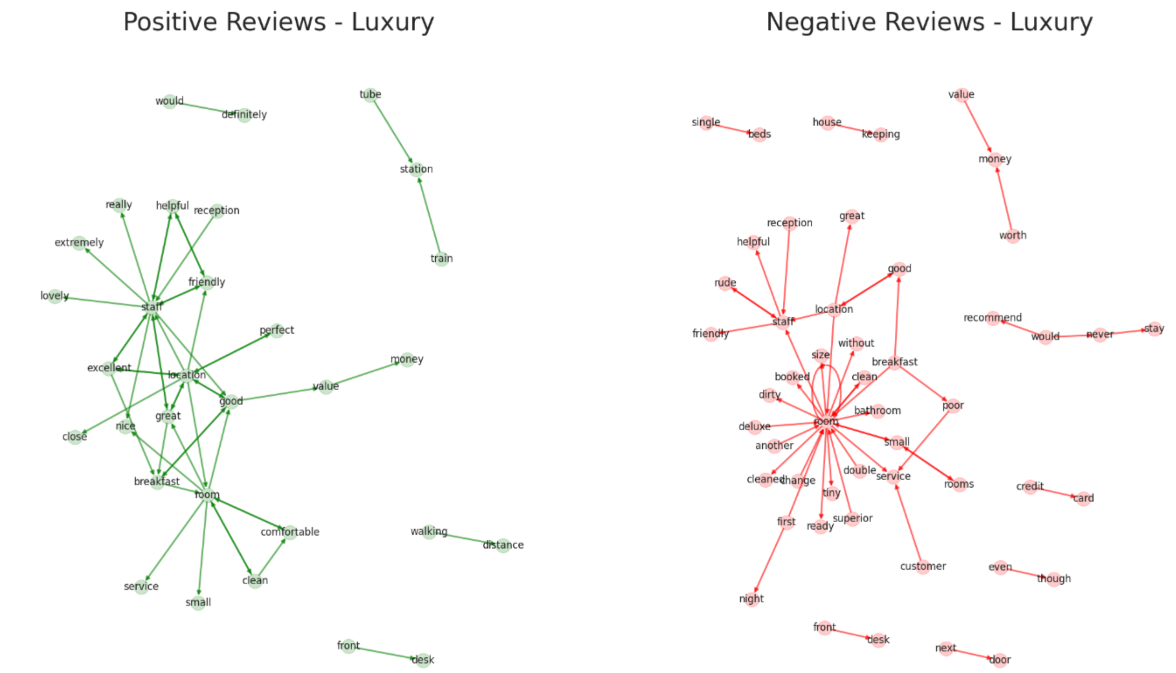

Figure 8 - Bigrams (luxury)

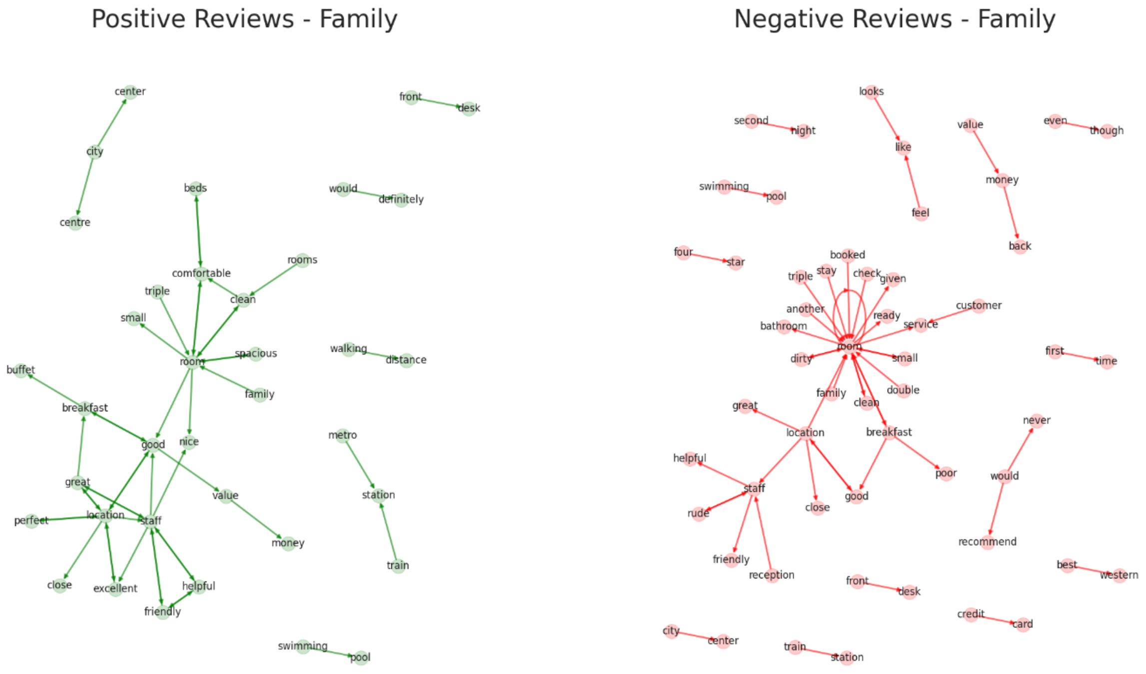

Figure 9 - Bigrams (Family)

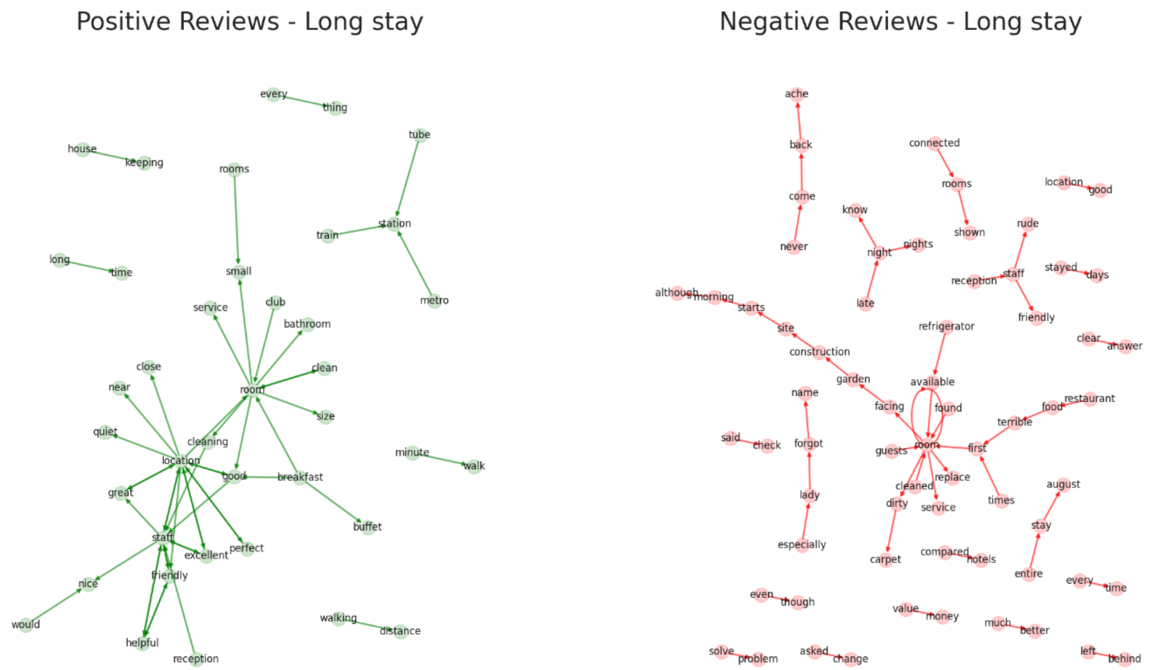

Figure 10 - Bigrams (long stay)

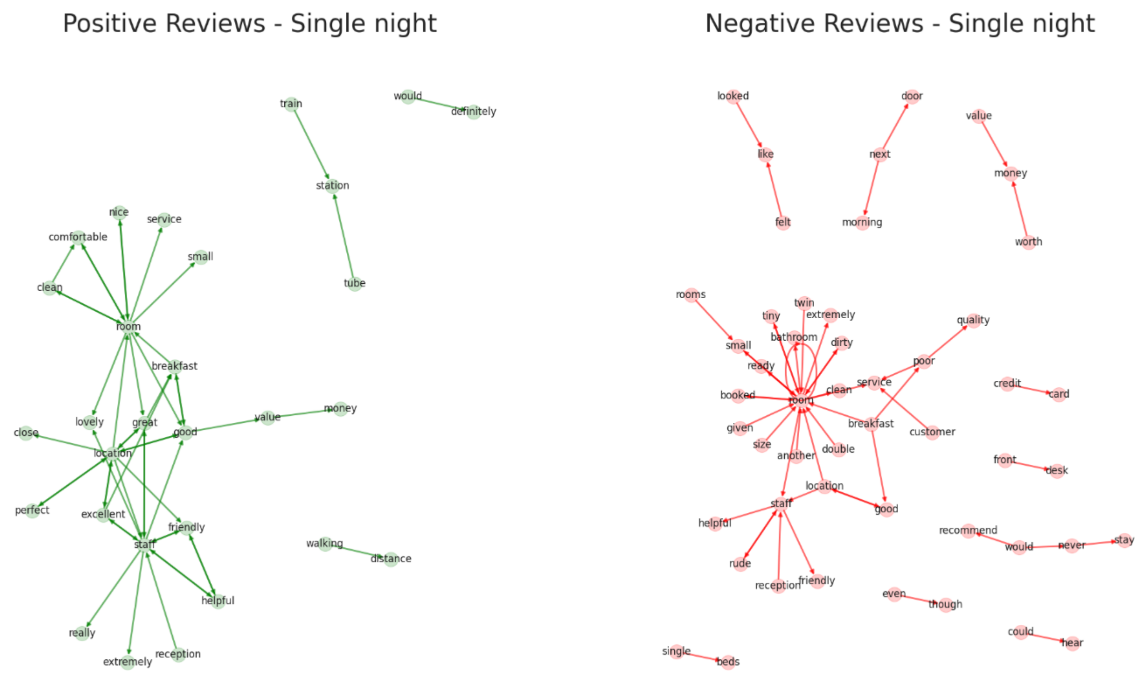

Figure 11 - Bigrams (Single night)

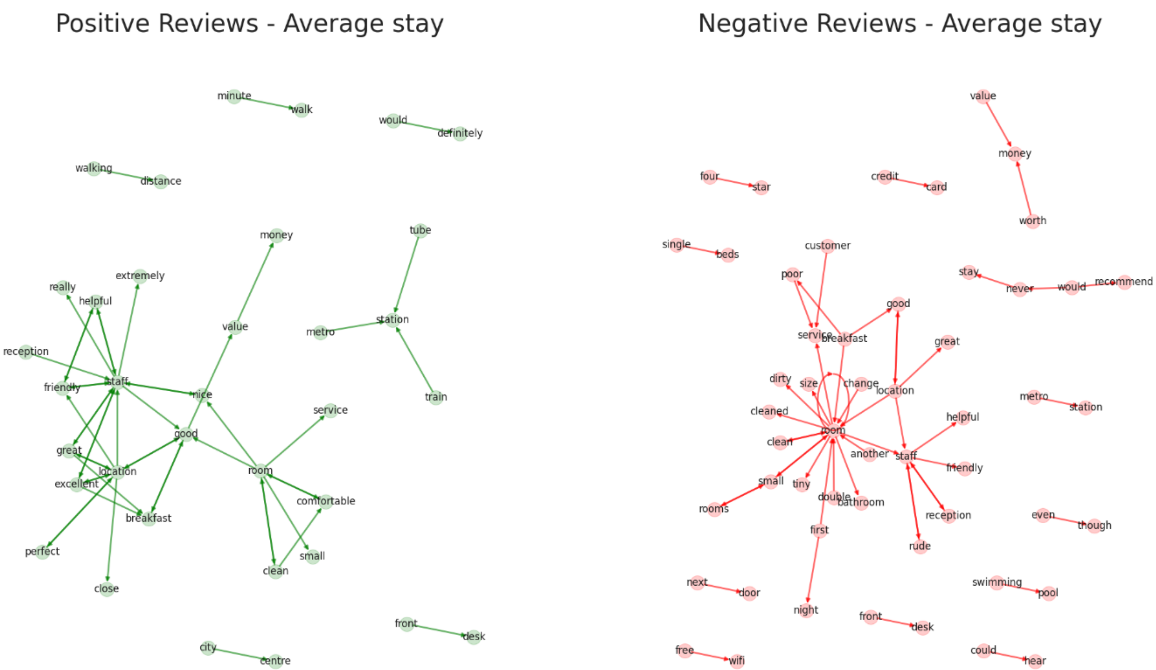

Figure 12 - Bigrams (average stay)

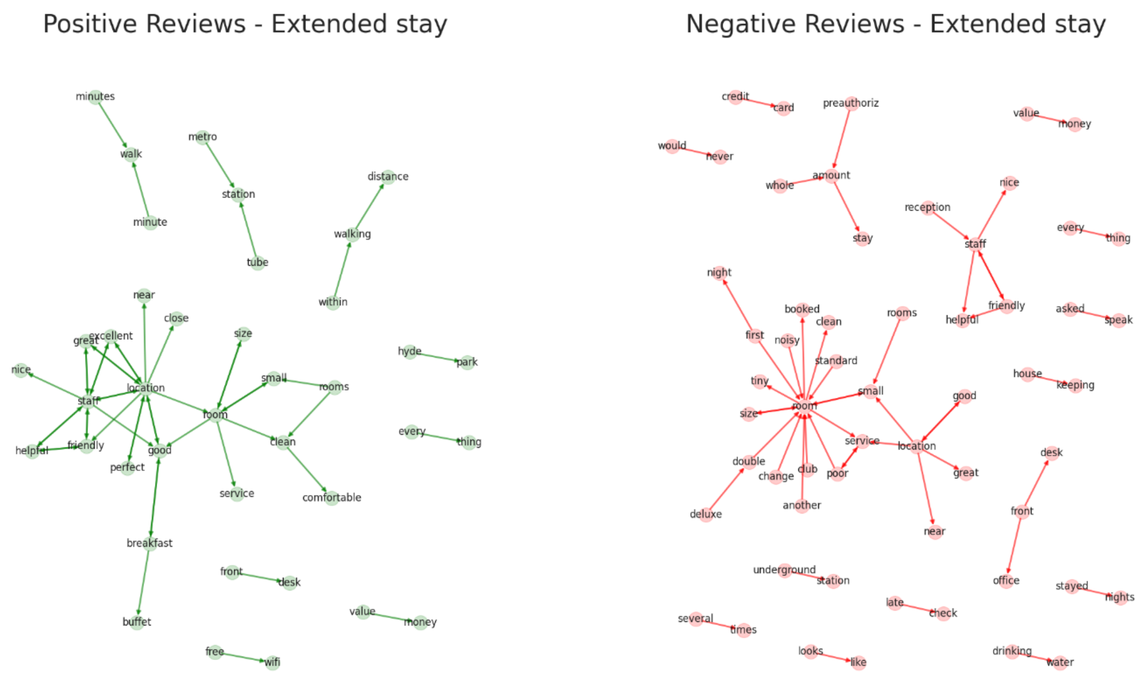

Figure 13 - Bigrams (extended stay)

References

Ahuja, R., Chug, A., Kohli, S., Gupta, S. and Ahuja, P., 2019. The impact of features extraction on the sentiment analysis. *Procedia Computer Science*, *152*, pp.341-348.

Asghar, M.Z., Khan, A., Ahmad, S. and Kundi, F.M., 2014. A review of feature extraction in sentiment analysis. *Journal of Basic and Applied Scientific Research*, *4*(3), pp.181-186

Aydin, N. and Birbil, S.I., 2018. Decomposition methods for dynamic room allocation in hotel revenue management. European journal of operational research, 271(1), pp.179-192.

Bürgisser, R. and Di Carlo, D., 2023. Blessing or Curse? The Rise of Tourism-Led Growth in Europe's Southern Periphery. JCMS: Journal of Common Market Studies, 61(1), pp.236-258.

Chang, V., Liu, L., Xu, Q., Li, T. and Hsu, C.H., 2023. An improved model for sentiment analysis on luxury hotel review. Expert Systems, 40(2), p.e12580.

De Pelsmacker, P., Van Tilburg, S. and Holthof, C., 2018. Digital marketing strategies, online reviews and hotel performance. *International Journal of Hospitality Management*, *72*, pp.47-55.

Dematis, I., Karapistoli, E. and Vakali, A., 2018. Fake review detection via exploitation of spam indicators and reviewer behavior characteristics. In SOFSEM 2018: Theory and Practice of Computer Science: 44th International Conference on Current Trends in Theory and Practice of Computer Science, Krems, Austria, January 29-February 2, 2018, Proceedings 44 (pp. 581-595). Springer International Publishing.

Duan, W., Gu, B. and Whinston, A.B., 2008. Do online reviews matter?—An empirical investigation of panel data. *Decision support systems*, *45*(4), pp.1007-1016

Elmogy, A.M., Tariq, U., Ammar, M. and Ibrahim, A., 2021. Fake reviews detection using supervised machine learning. *International Journal of Advanced Computer Science and Applications*, *12*(1).

Fontanarava, J., Pasi, G. and Viviani, M., 2017, October. Feature analysis for fake review detection through supervised classification. In *2017 IEEE international conference on data science and advanced Analytics (DSAA)* (pp. 658-666). IEEE.

Harish, B.S. and Revanasiddappa, M.B., 2017. A comprehensive survey on various feature selection methods to categorize text documents. *International Journal of Computer Applications*, *164*(8), pp.1-7.

He, S., Hollenbeck, B. and Proserpio, D., 2022. The market for fake reviews. *Marketing Science*, *41*(5), pp.896-921.

Hu, M. and Liu, B., 2004, August. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177).

Hu, N., Koh, N.S. and Reddy, S.K., 2014. Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision support systems*, *57*, pp.42-53.

Jing, L.P., Huang, H.K. and Shi, H.B., 2002, November. Improved feature selection approach TFIDF in text mining. In *Proceedings. International Conference on Machine Learning and Cybernetics* (Vol. 2, pp. 944-946). IEEE.

Kaplan, Andreas M., and Michael Haenlein. "Users of the world, unite! The challenges and opportunities of Social Media." *Business horizons* 53, no. 1 (2010): 59-68.

Lee, S.H. and Ro, H., 2016. The impact of online reviews on attitude changes: The differential effects of review attributes and consumer knowledge. *International Journal of Hospitality Management*, *56*, pp.1-9.

Mäntylä, M.V., Graziotin, D. and Kuutila, M., 2018. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, *27*, pp.16-32.

McHugh, M.L., 2013. The chi-square test of independence. Biochemia medica, 23(2), pp.143-149.

Mellinas, J.P., María-Dolores, S.M.M. and García, J.J.B., 2015. Booking. com: The unexpected scoring system. Tourism Management, 49, pp.72-74.

Mohawesh, R., Xu, S., Tran, S.N., Ollington, R., Springer, M., Jararweh, Y. and Maqsood, S., 2021. Fake reviews detection: A survey. *IEEE Access*, *9*, pp.65771-65802.

Ni, G., Feng, X. and Chen, L., 2020. Hotel room allocation with multiple sales channels: A perspective of minimax regret. Journal of Control and Decision, 7(4), pp.359-378.

Pang, B., Lee, L. and Vaithyanathan, S., 2002. Thumbs up? Sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.

Prabhat, A. and Khullar, V., 2017, January. Sentiment classification on big data using Naïve Bayes and logistic regression. In *2017 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-5). IEEE.

Qader, W.A., Ameen, M.M. and Ahmed, B.I., 2019, June. An overview of bag of words; importance, implementation, applications, and challenges. In *2019 international engineering conference (IEC)* (pp. 200-204). IEEE.

Qin, C., 2014, October. Room allocation for hotel revenue management considering consumer choice behavior. In 2014 International Conference on Management of e-Commerce and e-Government (pp. 323-328). IEEE.

Rana, R. and Singhal, R., 2015. Chi-square test and its application in hypothesis testing. Journal of the practice of cardiovascular sciences, 1(1), p.69.

Salminen, J., Kandpal, C., Kamel, A.M., Jung, S.G. and Jansen, B.J., 2022. Creating and detecting fake reviews of online products. *Journal of Retailing and Consumer Services*, *64*, p.102771.

Schuckert, M., Liu, X. and Law, R., 2015. Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing*, *32*(5), pp.608-621.

Sodanil, M., 2016. Multi-language sentiment analysis for hotel reviews. In *MATEC Web of Conferences* (Vol. 75, p. 03002). EDP Sciences.

Song, J., Parlar, M. and Yuan, Q., 2010, August. Decision Making of Hotel Room Allocation: A Statistic Game. In 2010 International Conference on Internet Technology and Applications (pp. 1-4). IEEE.

Sparks, B.A. and Browning, V., 2011. The impact of online reviews on hotel booking intentions and perception of trust. *Tourism management*, *32*(6), pp.1310-1323.

Tran, T., Ba, H. and Huynh, V.N., 2019. Measuring hotel review sentiment: An aspect-based sentiment analysis approach. In *Integrated Uncertainty in Knowledge Modelling and Decision Making: 7th International Symposium, IUKM 2019, Nara, Japan, March 27–29, 2019, Proceedings 7* (pp. 393-405). Springer International Publishing.

Vermeulen, I.E. and Seegers, D., 2009. Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism management*, *30*(1), pp.123-127.

Wang, N., Yang, J., Kong, X. and Gao, Y., 2022. A fake review identification framework considering the suspicion degree of reviews with time burst characteristics. *Expert Systems with Applications*, *190*, p.116207.

Zhai, Q., Tian, Y., Luo, J. and Zhou, J., 2023. Hotel overbooking based on no-show probability forecasts. Computers & Industrial Engineering, 180, p.109226.

Zhang, Z., Ye, Q., Law, R. and Li, Y., 2010. The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, *29*(4), pp.694-700.

Zhu, F. and Zhang, X., 2010. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing*, *74*(2), pp.133-148.