YRKESHÖGSKOLAN
NOVIA

Using Text Mining and Sentiment Analysis to Enhance Customer Experience and Service Quality in Hotel Operations: A Case Study of a Finnish Hotel

Suraj Shrestha

Degree Thesis

Thesis for a Master of Engineering (UAS) - degree

Degree Programme: Industrial Management and Engineering

Place and year: Vaasa 2024

DEGREE THESIS

Author:	Suraj Shrestha
Degree Programme and place of study:	Industrial Management and Engineering, M.Eng., Vaasa
Supervisor(s):	Ray Pörn, Novia UAS, Vaasa

Title: Using Text Mining and Sentiment Analysis to Enhance Customer Experience and Service Quality in Hotel Operations: A Case Study of a Finnish Hotel.

Date: 05.12.2024 Number of pages: 55 Appendices: 0

Abstract

This thesis explores the application of text mining and sentiment analysis to enhance customer experience and service quality in hotel operations, using a Finnish hotel as a case study. Leveraging internal customer feedback collected from November 2021 to October 2024, the study employs advanced analytical techniques, including Latent Dirichlet Allocation (LDA) for topic modeling and Bidirectional Encoder Representations from Transformers (BERT) for sentiment analysis. The research identifies key service aspects—such as breakfast experience, room and staff service, check-in experience, value for money, room comfort and noise, room amenities, overall experience and location, customer service quality—highlighting their impact on customer sentiment and satisfaction over time.

Findings reveal a dominant positive sentiment among customers but underscore recurring challenges in areas like noise control and breakfast variety. Aspect-based sentiment analysis pinpoints specific drivers of satisfaction and dissatisfaction, offering actionable insights for targeted improvements. The study concludes by recommending strategies for operational enhancements, emphasizing the role of data-driven decision-making in the hospitality industry. This work contributes to both academic discourse and practical management by showcasing the integration of natural language processing tools in improving guest experiences.

Language: English
Key Words: Text Mining, Sentiment Analysis, Customer Experience, Service Quality, Hospitality Management, Natural Language Processing, Aspect-Based Analysis

Table of Contents

1. Introduction

The hospitality industry is profoundly shaped by customer experience and service quality, both of which significantly influence customer satisfaction, brand loyalty, and financial performance. In an increasingly competitive market, hotels strive to differentiate themselves by offering superior customer experiences, which are largely influenced by customer feedback. Effective management of customer feedback enables hotels to address service issues, improve operations, and enhance guest satisfaction (Hay, 2024). Traditionally, customer feedback has been collected through surveys, comment cards, and direct interactions. However, with the rise of digital feedback channels, hotels now have access to vast amounts of unstructured data that can provide deep insights into customer perceptions and expectations (Ho, Withanage, & Khong, 2020).

In this context, this modern hotel located in Finland, provides an ideal setting for studying customer experience within an urban hotel environment. Since its opening, the hotel has gathered substantial internal customer feedback via online feedback forms. This study leverages text mining and sentiment analysis to analyze this feedback systematically. By extracting themes, sentiments, and trends from guest comments, this research aims to generate actionable insights for enhancing service quality.

Due to a confidentiality agreement with the hotel management, the specific name of the hotel has been anonymized throughout this document. This approach ensures the privacy and confidentiality of the hotel's information while allowing for a comprehensive analysis of the data collected.

1.1 Background of the Study

Customer experience (CX) and service quality have become critical components of business success in the hospitality industry, shaping guest satisfaction and loyalty (Maroco & Maroco, 2013). Service quality refers to the degree to which a service meets or exceeds customer expectations, often evaluated through dimensions such as reliability, responsiveness, assurance, empathy, and tangibles (Parasuraman, Zeithaml, & Berry, 1988). With increasing customer expectations, particularly among millennials and Generation Z, maintaining high standards of service quality is vital for hotels to stay competitive (Seyfi, Vo-Thanh, & Zaman, 2024).

Customer experience, on the other hand, is a broader concept that encompasses every interaction a guest has with a hotel, from booking to check-out. In recent years, there has been a paradigm shift in hospitality management toward a more customer-centric approach, where continuous improvement of CX is achieved by actively listening to and analyzing guest feedback (Dang & Nguyen, 2023). Traditional methods of analyzing customer feedback, however, have limitations. Survey responses and structured feedback forms often fail to capture nuanced customer opinions, while manual analysis of

unstructured feedback can be time-consuming and prone to bias (Kim & Kim, 2022). Consequently, text mining and sentiment analysis have emerged as powerful tools in hospitality management to process and analyze large volumes of unstructured feedback effectively.

1.2 Problem Statement

Despite the availability of vast amounts of customer feedback data, many hotels struggle to leverage this information effectively to improve service quality. Manual methods of analyzing feedback are often limited in scale and consistency, making it difficult to extract actionable insights across all aspects of hotel operations (Ray & Ma, 2021). Moreover, while positive feedback is generally easy to interpret, negative feedback often contains complex sentiment and requires nuanced analysis to identify specific areas of dissatisfaction. Without a systematic approach to analyzing customer feedback, valuable insights that could enhance guest satisfaction and operational efficiency may be overlooked.

This study addresses this gap by employing text mining and sentiment analysis to systematically analyze internal customer feedback at a modern hotel in Finland. By identifying themes and sentiments within guest comments, the research seeks to uncover key drivers of customer satisfaction and dissatisfaction, thus offering data-driven recommendations for enhancing the customer experience.

1.3 Objectives of the Study

The main objective of this study is to utilize text mining and sentiment analysis to systematically analyze internal customer feedback from a hotel in Finland, aiming to improve customer experience and service quality. Specific objectives include:

- **Objective 1**: To determine the overall sentiment of customer feedback and analyze trends over time.
- **Objective 2:** To identify the main topics that emerge from customer feedback at the hotel, assess how these topics influence customer sentiments, and analyze the trends in these topics and sentiments over time.
- **Objective 3**: To provide actionable recommendations based on the findings to enhance customer experience and improve service quality.

These objectives contribute to the broader goal of helping the hotel understand and respond more effectively to customer needs, thus fostering a positive guest experience.

1.4 Research Questions

This study seeks to answer the following research questions:

- **RQ1**: What is the overall sentiment of customer feedback, and how does it vary over time?
- **RQ2**: What are the main topics emerging from customer feedback at the hotel, how do they influence sentiments, and how do these topics and sentiments evolve over time?
- **RQ3**: What actionable recommendations can be derived from the findings to enhance customer experience?

These research questions provide a structured framework for analyzing feedback and translating insights into practical improvements in service quality.

1.5 Scope and Delimitations

The scope of this study is focused on analyzing internal customer feedback collected from the case-study hotel, spanning from its establishment through October 2024. The dataset consists of customer feedback gathered via email surveys, which are sent to guests after their stay to collect insights on various aspects of their experience, such as customer service, room quality, amenities, and overall satisfaction. While the feedback includes departure dates, response dates, name of the hotel, textual comments, Friends Level and numerical ratings, this study is delimited to analyzing only the response date and the textual comments provided in the surveys. The analysis will exclude departure dates, name of the hotel, Friends Level data, and the numerical ratings, focusing on text-based feedback and the timing of responses to uncover detailed themes and sentiments.

Additionally, this study employs computational processing through Google Colab for text analysis. In compliance with privacy and ethical considerations, all customer data will be anonymized to ensure adherence to GDPR guidelines (European Parliament and Council, 2016, Regulation (EU) 2016/679). These delimitations help to focus the analysis specifically on textual feedback and response timing, while upholding data protection standards and excluding variables that are outside the scope of this study.

1.6 Significance of the Study

This study contributes to both academic and practical fields by demonstrating how text mining and sentiment analysis can be effectively applied to analyze customer feedback in the hospitality industry. From an academic perspective, this research adds to the literature on customer experience management by exploring advanced data analysis techniques to identify guest sentiment patterns.

From a practical perspective, the insights generated in this study can help the case-study hotel address specific service issues, thereby enhancing customer satisfaction. For example, understanding customer feedback more deeply may enable the hotel to make targeted improvements in areas such as staff training, room amenities, and service responsiveness. This data-driven approach can lead to increased customer loyalty, higher guest retention, and improved operational efficiency.

Moreover, the findings from this study may serve as a model for other hotels seeking to leverage text mining and sentiment analysis as part of their customer feedback analysis process. In an industry where customer expectations are continuously evolving, adopting data-driven approaches to understand and respond to guest feedback is critical for long-term success (Troisi, Visvizi, & Grimaldi, 2023).

2. Literature Review

The growth of digital technology and data analytics has transformed how businesses in the hospitality industry understand and respond to customer needs. The industry has traditionally relied on guest feedback to gauge service quality and improve customer experience, with recent advancements in text mining and sentiment analysis providing a more systematic approach to analyzing large volumes of unstructured feedback (Mehraliyev, Chan, & Kirilenko, 2022). This chapter reviews the foundational concepts of customer experience and service quality in hospitality, examines text mining and sentiment analysis techniques, and highlights prior research on using these tools to enhance service quality in hotels.

2.1 Customer Experience and Service Quality in the Hospitality Industry

Customer experience (CX) and service quality have become central to the competitive strategies of hospitality businesses, which depend heavily on customer satisfaction and loyalty (Nyagadza, Mazuruse, Muposhi, & Chigora, 2022). Customer experience encompasses every interaction a guest has with a hotel, from booking to check-out, shaping their overall perception of the brand (Lemon & Verhoef, 2016). In the hospitality industry, CX directly impacts customer retention, word-of-mouth promotion, and revenue. Hotels aim to deliver exceptional experiences through attentive customer service, attractive amenities, and seamless operations, which are key components of positive guest experiences (Kim & Kim, 2022).

Service quality, a related but distinct concept, is often evaluated through dimensions like reliability, responsiveness, and assurance. One widely recognized model is the SERVQUAL framework, which measures customer perceptions across five dimensions: tangibles, reliability, responsiveness, assurance, and empathy (Parasuraman, Zeithaml, & Berry,

1988). Many studies indicate that service quality is a crucial determinant of customer satisfaction, which, in turn, influences repeat patronage and brand loyalty (Zygiaris, Hameed, Alsubaie, & Rehman, 2022).

However, with the rapid growth of digital feedback channels, the nature of customer feedback has evolved. Instead of structured survey responses, hotels now receive unstructured feedback through online reviews and social media, making it essential to adopt data-driven methods for efficient analysis (Zarezadeh, Rastegar, & Xiang, 2022). Thus, modern hotels are increasingly turning to text mining and sentiment analysis as tools to gain actionable insights from customer feedback.

2.2 Text Mining Techniques in Hospitality

Text mining refers to the process of extracting meaningful information from unstructured text data. In hospitality, text mining is used to uncover patterns, trends, and themes in customer feedback (Bi, Zhu, & Han, 2024). Key techniques include tokenization, which splits text into individual words or terms, and lemmatization, which reduces words to their base forms for easier analysis (Bird, Klein, & Loper, 2009). These methods enable researchers to preprocess data, making it suitable for more complex analysis.

One commonly applied text mining technique in hospitality research is topic modeling, a machine learning approach used to identify latent themes in text data. A popular algorithm for topic modeling is Latent Dirichlet Allocation (LDA), which categorizes text into different topics based on word co-occurrence patterns (Blei, Ng, & Jordan, 2003). In a hotel context, topic modeling can reveal common themes in guest feedback, such as comments on room cleanliness, staff professionalism, or facility quality, which help hotels prioritize improvements (Sutherland, Sim, Lee, Byun, & Kiatkawsin, 2020).

2.3 Sentiment Analysis: Concepts and Applications

Sentiment analysis, often referred to as opinion mining, is a technique used to determine the emotional tone of text data. In hospitality, sentiment analysis enables hotels to classify guest feedback as positive, negative, or neutral, providing a general sense of customer satisfaction (Mehraliyev, Chan, & Kirilenko, 2022). Sentiment analysis can be rule-based, where specific words and phrases are associated with positive or negative sentiments, or machine-learning-based, where models are trained on labeled data to predict sentiment.

In recent years, sentiment analysis has advanced through natural language processing (NLP) techniques such as transformer-based models like BERT (Bidirectional Encoder Representations from Transformers) and VADER (Valence Aware Dictionary and sEntiment Reasoner), which are highly effective in capturing nuanced sentiment from text (Samy Talaat, 2023; Hoti & Ajdari, 2023). For instance, in a study on hotel reviews, BERT-based sentiment analysis could accurately identify customer sentiment across various aspects of service quality, such as room comfort and staff interactions.

Aspect-based sentiment analysis (ABSA) is another extension of sentiment analysis specifically valuable for hotel feedback, as it categorizes sentiment based on specific aspects of service, such as "cleanliness" or "location" (Samodra & Sutoyo, 2023). ABSA enables hotels to understand sentiment at a more granular level, which aids in identifying specific areas that contribute to customer satisfaction or dissatisfaction. For example, if sentiment is consistently negative for "wifi quality," the hotel can focus on this area for improvement.

2.4 Topic Modeling and Aspect-Based Sentiment Analysis

Topic modeling and aspect-based sentiment analysis (ABSA) are powerful tools for understanding the structure of customer feedback. While topic modeling uncovers major themes in text data, ABSA dives deeper into the sentiment associated with each identified theme (Dang & Nguyen, 2024). For instance, a topic modeling approach might reveal that guests frequently mention "staff friendliness" and "room cleanliness." ABSA can then analyze whether these mentions are associated with positive or negative sentiment.

In hospitality research, combining topic modeling with ABSA has proven effective in drawing detailed insights from customer feedback. Using both techniques on hotel reviews could help hotels address specific issues, such as improving "check-in efficiency" when customer sentiment on this topic is negative. This integration enables a more targeted approach to service quality improvements and enhances the hotel’s ability to act on feedback in real-time (Ounacer et al., 2023).

2.5 Prior Studies on Customer Feedback Analysis in Hospitality Sector

A substantial body of research has applied text mining and sentiment analysis to better understand customer feedback in the hospitality sector. These studies highlight the growing importance of these analytical techniques in enhancing service quality and improving customer satisfaction. Below is a detailed review of such significant studies, starting from older research and progressing to more recent contributions.

The pioneering study by Parasuraman et al. (1988) introduced the SERVQUAL model, which laid the foundation for understanding service quality in the hospitality sector. Their model provided five key dimensions—tangibles, reliability, responsiveness, assurance, and empathy—which helped hotels assess customer satisfaction. Although not directly focused on text mining, this study influenced later research into customer feedback analysis, as it provided a framework for evaluating the core aspects of hotel service quality.

Chaudhuri and Holbrook (2001) explored drivers of customer loyalty and satisfaction that are fundamental in understanding customer feedback. Although not directly focused on sentiment analysis, it highlighted how trust and emotional response affect brand perception—a basis for analyzing hotel reviews today. The study has influenced subsequent

sentiment analysis techniques by identifying key factors in customer satisfaction, setting the stage for feedback analysis in hospitality.

Hu and Liu (2004) introduced an aspect-based approach to analyzing reviews by extracting features and associating positive or negative sentiments with each. Although not specific to hotels, it has shaped how hospitality reviews analyze specific service aspects, such as room quality or staff behavior. This model set a foundation for sentiment analysis in hospitality by isolating actionable feedback points.

Pang and Lee (2008) comprehensively covered sentiment analysis models, from rule-based systems to machine learning, which have become integral in hotel review analysis. Their discussion of algorithms, particularly SVM and Naïve Bayes, has influenced many applications in the hospitality sector, where customer feedback analysis seeks accuracy in classifying positive and negative sentiments.

Zhang, Ye, Law, and Li (2010) investigated the effects of electronic word-of-mouth (eWOM) on online restaurant popularity by analyzing consumer and editor reviews. This study found that both review types significantly influence a restaurant's online visibility, but consumer reviews had a particularly strong impact on reputation. The research highlighted the importance of positive eWOM in boosting online popularity, which can be instrumental for hospitality managers aiming to enhance brand perception and attract more customers. By analyzing the sentiment in customer feedback, restaurant managers can better understand guest experiences and identify areas for improvement. This study serves as a valuable foundation for examining the role of online reviews in broader hospitality settings, including hotels, underscoring how eWOM shapes business outcomes in customer-centric industries.

Dominici and Guzzo (2010) conducted a case study on customer satisfaction in the hotel industry in Sicily, focusing on the relationship between service quality, customer satisfaction, and loyalty. Their findings highlight that service quality is the key determinant of customer satisfaction, which in turn influences customer loyalty. They argue that satisfied customers are more likely to return and recommend the hotel to others, contributing to improved profitability through repeat business and word-of-mouth referrals. The study is valuable for understanding the practical implications of customer satisfaction in a competitive hotel market, and it underscores the importance of meeting customer expectations to foster long-term loyalty.

Vinodhini & Chandrasekaran (2012) provided a comprehensive review of sentiment analysis methods used in various domains, including hospitality. Vinodhini and Chandrasekaran highlighted the primary challenges in sentiment extraction, such as sarcasm and spam filtering, influencing subsequent studies in text mining for customer feedback in hotels. They emphasized the need for hybrid approaches to improve sentiment classification accuracy.

Mauri & Minazzi (2013) focused on how online reviews shape customer expectations and booking decisions. Using sentiment analysis, the authors found that positive reviews led to higher customer expectations, while negative feedback deterred bookings. This study

highlighted the importance of review monitoring to manage customer expectations effectively.

Blal and Sturman (2014) explored the effects of online reviews on hotel room sales, focusing on both the quality (ratings) and quantity (number of reviews) of feedback. Their research demonstrated that high-quality reviews, particularly for upscale hotels, were strongly correlated with increased room sales. On the other hand, the quantity of reviews played a more significant role in improving sales for midscale and economy hotels. This study highlighted the financial implications of customer feedback, underscoring the importance of not just accumulating reviews but also ensuring their quality. These findings suggest that sentiment analysis tools can be instrumental in identifying which reviews have the greatest potential to impact hotel performance, offering a practical application of customer feedback management in the hospitality sector.

Xiang, Schwartz, Gerdes, & Uysal (2015) utilized big data and sentiment analysis on TripAdvisor reviews to identify key drivers of satisfaction in the hotel industry. The researchers found that cleanliness and staff service were most correlated with high satisfaction, while negative sentiment clustered around noise issues and outdated amenities. Their use of big data underscored the increasing importance of automated text analytics in extracting actionable insights for hospitality management.

Kwok and Xie (2016) examined the factors influencing the helpfulness of online hotel reviews and highlighted the role of management responses. Their study found that management's timely and personalized responses to customer reviews significantly increased the perceived helpfulness of those reviews. This finding emphasizes the importance of active review management, suggesting that when hotels respond thoughtfully to guest feedback, it not only resolves customer concerns but also enhances the value of the reviews for future customers. This study adds to the growing body of research showing how effective management responses can improve a hotel's online reputation.

Yu (2016) explores various supervised machine learning methods for sentiment analysis on TripAdvisor hotel reviews. The study focuses on not only overall opinions but also aspect-based opinions, including service, rooms, location, value, cleanliness, sleep quality, and business service. The results showed that the accuracy of the predictors reached 70% to 75% for star-rating and about 85% to 90% for polarity.

Liu et al. (2017) investigated how big data analytics can be applied to analyze user-generated reviews at scale to uncover language-specific drivers of hotel satisfaction. By examining a massive dataset of 412,784 reviews, they used sentiment analysis to identify factors that most strongly influence customer satisfaction in different languages. Their study demonstrated the potential of big data techniques in the hospitality industry to provide granular insights into customer feedback, allowing hotels to tailor their service offerings more effectively to diverse customer groups. This research aligns with the growing importance of big data in enhancing the precision of customer satisfaction analysis in the hotel sector.

Tuerlan, Li, and Scott (2021) provide a comprehensive review of emotion research in hospitality and tourism, synthesizing 178 studies published from 2004 to 2019. They critique existing literature for oversimplifying customer emotions as merely positive or negative, without accounting for the complexities of discrete emotions. The authors suggest that most emotion studies in the field rely heavily on psychological frameworks, which may not be entirely suited to the hospitality context. This gap highlights the need for more context-specific emotion scales and advanced methodologies to more accurately capture customer emotions during service encounters. The review also emphasizes the importance of emotion measurement in understanding customer experiences, an area closely tied to sentiment analysis and its role in improving service quality and customer satisfaction in the hospitality sector. By drawing on appraisal theories of emotions, the authors propose a more nuanced approach to understanding the role of emotions in shaping customer perceptions and behaviors.

Iqbal et al. (2022) explored the application of deep learning techniques in sentiment analysis of consumer reviews. The study focuses on leveraging advanced machine learning methods to extract meaningful insights from large volumes of customer feedback. By using deep learning models, the authors demonstrate how such techniques can significantly improve the accuracy and efficiency of sentiment classification, which is crucial for understanding customer sentiment in the hospitality industry. Their work contributes to the evolving field of sentiment analysis, emphasizing the power of AI in enhancing customer feedback analysis.

Tang, Wang, and Kim (2022) investigated the role of customer reviews in predicting conversion rates for online hotel bookings. The study highlights how reviews, by influencing potential customers' decisions, can significantly affect booking outcomes. Using sentiment analysis and other metrics, the authors identified key factors in customer feedback that correlate with higher conversion rates, offering valuable insights for hotel managers to optimize their online presence and marketing strategies. This research contributes to the growing body of work on the predictive power of customer reviews in the hospitality industry.

Puh and Bagić Babac (2022) explores the use of machine learning and deep learning models to predict sentiment and ratings from tourist reviews. The authors used models such as Naïve Bayes, Support Vector Machines (SVM), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Bidirectional LSTM (BiLSTM) to classify reviews into positive, negative, or neutral sentiments, and to predict ratings from one to five stars. The data for training these models was collected from TripAdvisor. The results showed that deep learning models, particularly BiLSTM, achieved higher accuracy compared to traditional machine learning models. The study highlights the practical implications of these models in improving customer experience, forecasting tourist arrivals, and enhancing marketing strategies.

Mehraliyev, Chan, & Kirilenko (2022) conducts a systematic review and critically analyzes the sentiment analysis literature in hospitality and tourism from methodological (data sets and analyzes) and thematic (topics, theories, key constructs and their relationships) perspectives. The authors reviewed 70 papers from hospitality and tourism categories of

Web of Science and Scopus databases. They identified 5 topics and 27 sub-topics, with the major theme being market intelligence. The study highlights the need for more theoretical frameworks and big data analysis in sentiment analysis research.

Malik and Bilal (2024) offer a comprehensive survey on the application of Natural Language Processing (NLP) techniques in analyzing online customer reviews. Their study provides a taxonomy for classifying NLP methods and discusses the open research challenges in this field. The authors explore the importance of sentiment analysis, aspect-based analysis, and opinion mining in understanding customer feedback, emphasizing the need for advanced NLP approaches to tackle issues such as sarcasm and multilingual data in customer reviews. This study provides valuable insights into the evolving landscape of customer feedback analysis in the hospitality sector.

Adikari, Nguyen, Nawaratne, De Silva, and Alahakoon (2024) addresses the growing importance of online hotel review platforms and the need for decision-makers in the hospitality sector to extract valuable information from these vast sources. The authors propose an approach that uses Natural Language Processing (NLP) techniques to convert unstructured textual reviews into a structured representation of emotions and hotel aspects. They conducted a segmentation analysis to identify distinct emotion and concern-based profiles of customers and hotels using a self-organizing unsupervised algorithm. The study demonstrated the practicality of this approach using 22,450 online reviews from 44 hotels. The insights from emotion analysis and review segmentation help develop targeted customer management strategies and informed decision-making.

Ounacer, Daif, El Ghazouani, & Azzouazi, (2024) explores the application of sentiment analysis in the hospitality industry, focusing on how AI and machine learning techniques can be used to analyze hotel reviews. The authors discuss the use of traditional machine learning techniques such as Support Vector Machines (SVM), Naïve Bayes, and Logistic Regression, as well as deep learning models like Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and BERT (Bidirectional Encoder Representations from Transformers). The study highlights that BERT often surpasses traditional methods in both precision and contextual understanding. The findings suggest that sentiment analysis can significantly enhance hotel services by providing insights into customer emotions and preferences, ultimately leading to improved customer satisfaction and business success.

These studies highlight the evolution of sentiment analysis and text mining in the hotel industry, from foundational methodologies to advanced AI applications. Together, they reveal the critical role of customer feedback analysis in enhancing service quality, improving customer satisfaction, and supporting data-driven decision-making in hospitality.

2.6 Conceptual Framework

This study's conceptual framework integrates the components of customer feedback analysis—text mining, sentiment analysis, and customer experience improvement—into a structured approach. Customer feedback serves as the foundation of this framework, and

is processed using text mining techniques to identify key themes and aspects of service quality. Sentiment analysis is then applied to these themes to gauge overall satisfaction levels and identify specific areas of concern.

The framework is grounded in prior research that demonstrates how analyzing customer feedback using data-driven techniques can enhance service quality. By following this framework, this study aims to provide the hotel under study with actionable insights into customer perceptions and recommend strategies for improving customer experience.

3. Research Methodology

This chapter outlines the research methodology used in this study. The primary aim of this research is to gain insights from guest feedback to identify key service quality dimensions and areas for improvement. The analysis leverages text mining and sentiment analysis techniques to process unstructured feedback data, providing a comprehensive view of customer satisfaction from November 2021 till October 2024.

3.1 Research Approach and Research Design

Research Approach

The research approach employed in this study is predominantly **quantitative**, utilizing text mining and sentiment analysis to analyze internal customer feedback. While the primary data source consists of unstructured qualitative feedback, the objective is to derive measurable and actionable insights. Text mining technique such as topic modeling, and sentiment analysis transform this textual data into quantifiable information, enabling a systematic analysis of trends, patterns, and sentiment distributions over time.

Customer feedback is classified into sentiment categories (positive, negative, or neutral) and analyzed to identify recurring themes. This classification converts qualitative responses into structured data, facilitating a quantitative interpretation of the qualitative content. This approach not only captures the emotional tone and thematic focus of the feedback but also provides a data-driven foundation for decision-making and trend analysis.

Research Design

This study employs an **exploratory** and **descriptive** research design to uncover patterns and understand the sentiments expressed by hotel guests through internal feedback. The **exploratory** aspect allows for an in-depth examination of the unstructured feedback, revealing unexpected insights into customer preferences and perceptions. As this is among the first studies to apply text mining to the hotel's internal feedback, an exploratory design is particularly suitable (Huang, Liang, & Choi, 2022).

The **descriptive** component complements the exploratory approach by systematically quantifying and categorizing the feedback. This involves summarizing the data to identify trends, sentiments, and themes related to the hotel's service offerings. Together, these approaches provide a comprehensive understanding of guest experiences and generate actionable insights for hotel management (Buehring & O'Mahony, 2019).

3.2 Data Collection

3.2.1 Source of Data

The data used in this study consists of internal customer feedback collected through the hotel's routine customer satisfaction surveys. These surveys include both quantitative ratings and qualitative open-ended responses. For this research, only the unstructured text from the qualitative responses was analyzed.

3.2.2 Data Timeframe and Data Extraction Process

The feedback data spans from November 14, 2021, to October 31, 2024, providing a robust dataset covering the early stages of the hotel's operations and recent trends in guest satisfaction.

Feedback responses were retrieved from the hotel's internal database, ensuring that only anonymized entries were used to protect guest privacy. Any personally identifiable information (PII) was excluded during the data extraction process. The retrieved responses were then formatted into a structured dataset, which included customer comments along with basic metadata such as the date of response. This finalized dataset was used for further processing and analysis.

3.2.3 Statistical Overview of the Dataset

To ensure the integrity and quality of the dataset obtained from the hotel, an initial statistical analysis was conducted. This process provided a foundational understanding of the data and allowed for the identification of potential anomalies or issues that could influence subsequent stages of analysis. The statistical overview focused on three primary aspects: **counting blank comments**, **word count statistics**, and **language detection**.

The first step in the analysis involved identifying and quantifying blank or empty comments within the dataset. These entries, which lack substantive content, are typically either the result of system errors, customer omissions, or irrelevant data capture. By counting the number of blank comments, it was possible to assess the proportion of the dataset that was potentially uninformative and make decisions regarding their exclusion or treatment.

Word count statistics were computed for all non-blank entries to evaluate the richness and variability of the customer feedback. Metrics such as the minimum, maximum, average, and median word counts were derived. These statistics provided insights into the level of detail in customer reviews, helping to identify trends such as overly brief responses that may lack depth or unusually verbose comments that might require special handling.

Given the potential diversity of the customer base, the dataset was analyzed to detect the language of each comment. Language detection was critical in determining the linguistic distribution of the dataset, identifying non-target language entries, and deciding whether translation or filtering was necessary.

3.3 Data Preprocessing

Data preprocessing is a critical step in preparing the unstructured feedback for analysis. This process involves various techniques to clean, normalize, and transform the text data into a format suitable for text mining and sentiment analysis (Liu, 2012).

3.3.1 Removal of Blank Comments and Noisy Comments

To prepare the dataset for meaningful analysis, a data cleaning process was undertaken to remove entries that lacked substantive content or were otherwise unsuitable for text-mining and sentiment analysis. This step focused on eliminating blank comments and noisy comments that could undermine the quality and reliability of the analysis.

Removal of Blank Comments

All rows in the dataset where comments were completely empty or consisted only of whitespace were identified and removed. These entries lacked any textual content and, therefore, did not contribute value to the research objectives.

Handling Noisy Comments

During the language detection process, 227 feedback entries were categorized as "unknown language." A manual inspection of these entries revealed that they contained various forms of noise, such as:

- Symbols,
- Punctuation marks without accompanying text,
- Combinations of symbols and numbers.

As these noisy comments held no semantic meaning and were irrelevant for text-mining or sentiment analysis, they were excluded from the dataset (Ved, 2020). Similarly, rows containing only numbers or combinations of numbers, symbols, and spaces were also removed, as they did not provide useful insights for the analysis.

By removing these blank and noisy comments, the dataset was refined to include only entries with meaningful textual content, ensuring a higher quality of data for subsequent analytical processes. This step was critical to maintaining the focus and validity of the study's findings.

3.3.2 Translation of Multilingual Feedback

The hotel's diverse, international clientele resulted in customer feedback being provided in multiple languages, primarily Finnish, English, and Swedish. To ensure consistency and reliability in analysis, it was necessary to standardize the dataset by translating all non-English comments into English.

To achieve this, the Python library Googletrans (version 4.0.0-rc1), which interfaces with Google Translate's API, was employed. The translation process was implemented using a script executed in Google Colab, enabling seamless integration with the dataset. The tool automatically detected the language of each comment and translated it into English, eliminating the need for manual language identification.

This translation step was crucial for achieving uniformity across the textual data, enabling effective downstream processing such as text-mining and sentiment analysis. By consolidating the dataset into a single language, potential inconsistencies due to linguistic variations were mitigated, ensuring that the analysis could accurately reflect customer sentiments and feedback.

3.4 Overall Sentiment and Trends in Customer Feedback

To explore the overall sentiment of the collected feedback data and identify prevailing trends, a comprehensive sentiment analysis was conducted. This section outlines the methodology adopted for sentiment assessment and the extraction of temporal trends to understand the dynamics of feedback over time.

The analysis involved the following steps:

1. **Data Preprocessing**: Feedback data, that was translated to English, was prepared for analysis through basic cleaning steps, such as:
    - Removing special characters, URLs and extraneous spaces.
    - Ensuring consistent text formatting (e.g., lowercase conversion).
2. **Sentiment Scoring**: The VADER (Valence Aware Dictionary for Sentiment Reasoning) model, a lexicon-based approach well-suited for sentiment analysis of short texts such as customer feedback, was utilized. This model accounts for both intensity and sentiment polarity, making it ideal for analyzing customer feedback (Hutto & Gilbert, 2014). Each feedback entry was assigned a sentiment score and categorized as:

- Positive
- Neutral
- Negative

3. **Visualization and Trend Analysis**: To visualize overall sentiment and to identify trends in sentiment over time:
    - Pie charts were employed to illustrate the overall distribution of sentiment categories in the dataset.
    - Line graphs were used to display changes in sentiment across specific intervals covering the period from November 2021 to October 2024.

3.5 Text Mining and Sentiment Analysis Process

This subchapter outlines the comprehensive approach used to preprocess textual data, perform topic modeling, and conduct sentiment analysis. The process ensures strong data handling, from cleaning raw text to visualizing trends, which provides valuable insights into internal customer feedback.

3.5.1 Data Acquisition and Preparation

Data Upload and Structure

The customer feedback data, that was translated to English, was uploaded and imported into the analysis environment. The dataset, typically in Excel format, included essential columns such as comments, and response dates.

Initial Exploration

A preliminary inspection of the dataset was conducted to understand the data structure and identify potential issues such as missing values or inconsistent entries. Displaying a sample of the comments provided an overview of the feedback's content.

3.5.2 Text Preprocessing and Cleaning

Text Normalization

To prepare the data for analysis, all textual content was standardized through several preprocessing steps:

- Character Filtering: Special characters, numbers, and unnecessary spaces were removed, ensuring only alphabetic characters remained.
- Case Normalization: All text was converted to lowercase to maintain consistency.

Tokenization and Lemmatization

Each comment was tokenized, splitting the text into individual words. Tokenization allows for focused analysis on specific components of the text (Bird et al., 2009).

Subsequently, lemmatization was applied to reduce each word to its base or root form. This step improves the accuracy of the analysis by reducing redundancy in the data (Tan et al., 2023). Part-of-speech (POS) tagging helped ensure that words were lemmatized accurately according to their context (e.g., nouns, verbs, adjectives).

Stop-word Removal

Common English stop-words (e.g., "the," "and," "is") which do not carry significant meaning were removed to focus the analysis on meaningful content (Pang & Lee, 2008). This step was crucial for topic modeling, as stop-words can obscure the core themes in the data.

3.5.3 Topic Modeling Using Latent Dirichlet Allocation (LDA)

Preparation for LDA

The preprocessed text was vectorized using the CountVectorizer method. This transformation converted the cleaned comments into a document-term matrix, capturing word frequencies necessary for topic modeling.

Optimal Topic Selection

To determine the optimal number of topics, the LDA model was run iteratively with varying topic numbers (ranging from 1 to 10). For each run, a coherence score was calculated. Coherence measures how semantically interpretable the topics are, with higher scores indicating better performance (Zvornicanin, 2024). The number of topics with the highest coherence score was selected for the final model.

Topic Extraction

Once the optimal number of topics was identified, the LDA model was applied to the dataset. LDA groups words that frequently appear together and categorizes them into topics, revealing various themes (Blei et al., 2003). Each topic was characterized by its most significant words (typically the top 10), providing an initial understanding of the themes.

Visual Representation and Interpretation

Word clouds were generated for each topic to visualize the most frequent words, enhancing the interpretability of the extracted themes. This visualization helped identify the core concepts within each topic at a glance. Then these topics were manually reviewed and categorized into more interpretable labels (e.g., "Breakfast Experience," "Room Comfort and Noise").

3.5.4 Sentiment Analysis Using BERT

Model Selection and Initialization

Sentiment analysis was conducted using a pre-trained BERT (Bidirectional Encoder Representations from Transformers) model designed for multilingual sentiment classification. BERT is a deep learning model that excels in capturing the nuances of human language, including sentiment. It has been shown to perform well on sentiment analysis tasks, especially when fine-tuned for domain-specific datasets (Devlin et al., 2019).

Sentiment Classification

Each cleaned comment was processed through the BERT model, and the sentiment label was assigned based on the highest probability. This model categorized text into five sentiment classes: strongly negative (1 star), negative (2 stars), neutral (3 stars), positive (4 stars) and strongly positive (5 stars). This step translated qualitative feedback into measurable sentiment scores.

3.5.5 Aspect-Based Sentiment Analysis (ABSA)

Topic-Specific Sentiment Aggregation

To understand sentiment at the aspect level, the feedback was grouped according to the assigned topics. The sentiment distribution within each topic was calculated, showing the proportion of positive, neutral, and negative sentiments. By examining how sentiment varies across different aspects, hotel management can identify areas that require attention (Zhang, Li, Deng, Bing, & Lam, 2022). This approach provided a granular view of customer perceptions across different service aspects.

Visualization of Sentiment Distribution

Bar charts were created to display sentiment distributions across topics. Each bar represented the number of comments within a topic, segmented by sentiment class. This visualization highlighted areas with strong positive or negative feedback, guiding targeted improvements.

3.5.6 Temporal Analysis of Sentiment and Topics

Trends Over Time

To analyze changes in customer feedback over time, the response dates were extracted, and data was grouped by month. Sentiment trends were visualized for each topic, showing how customer perceptions evolved. Line charts illustrated the fluctuations in sentiment classes across different periods, revealing patterns such as seasonal variations or responses to service changes.

Topic Frequency Analysis

Similarly, the frequency of each topic over time was analyzed. Line plots demonstrated how the prominence of various themes shifted, indicating evolving customer concerns or interests.

3.6 Tools and Software

This section outlines the tools, libraries, and software platforms utilized to implement the text mining and sentiment analysis methods detailed in this study. The workflow was primarily executed using Google Colab, utilizing Python's robust ecosystem of data science libraries.

3.6.1 Development Environment

- Google Colab: A cloud-based platform providing an interactive Python environment with GPU support, facilitating efficient execution of machine learning models. It enabled seamless collaboration, data storage, and resource utilization.

3.6.2 Libraries and Packages

Data Handling and Manipulation

- Openpyxl: For reading and writing Excel files with the Colab environment, enhancing compatibility with Excel-based data sources.
- Pandas: For data loading, cleaning, manipulation, and exporting.
- NumPy: For numerical operations and handling arrays.

Text Processing

- Langdetect: Used to identify the language of each feedback comment. This library facilitated multilingual analysis by detecting and categorizing comments based on language.
- String Module: Used to handle operations involving punctuation and symbols.
- nltk: Used for natural language processing tasks such as stop-word removal, tokenization, and lemmatization (Bird et al., 2009).
- re (Regular Expressions): For text cleaning, such as removing special characters and extra spaces.

Text Translation

- Googletrans (Google Translate API):

- ✓ The translation of non-English text into English was performed using the *googletrans* library, which provides an interface to Google Translate. This tool was essential in standardizing the language of the dataset, ensuring that all textual feedback could be analyzed in a consistent language.
- ✓ A specific version of the library, googletrans==4.0.0-rc1, was installed to ensure compatibility and stability.
- ✓ The translation process was automated using a custom Python function, which applied the translator to each comment in the dataset. This function utilized a progress tracker (*tqdm*) to monitor translation progress.

Topic Modeling

- scikit-learn:
  - ✓ LatentDirichletAllocation (LDA): Implemented for extracting latent topics from the text data.
  - ✓ CountVectorizer: Converts text data into a matrix of token counts for topic modeling.
- genism:
  - ✓ CoherenceModel: Evaluates the quality of topics using coherence scores.
  - ✓ Dictionary: Converts tokenized documents into a bag-of-words format.

Sentiment Analysis

- NLTK (Natural Language Toolkit):
  - ✓ The nltk library, specifically the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis tool, was used to classify feedback into positive, negative, or neutral sentiments.
- Transformers Library (Hugging Face):
  - ✓ BERT model: Hugging Face's transformers library is used to implement BERT for sentiment analysis (Wolf et al., 2020).
  - ✓ Pipeline API: Simplifies applying the pre-trained model for text classification.

Visualization

- Matplotlib and Seaborn: For generating visualizations, including line charts, histograms, pie charts, and sentiment trends and topic distributions over time.
- WordCloud: To visualize top words in each topic.

3.6.3 Additional Tools

- Excel Handling: Data was imported and exported in Excel format using openpyxl and Pandas.
- File Management: Utilized Google Colab’s files module for uploading and downloading datasets and processed results.

3.7 Workflow Execution Overview

For easy in execution the total work was divided into five major parts and executed one after another. These are described below:

1. Statistical Analysis of Dataset

i. Data Loading: The *pandas* library facilitated the import and initial inspection of the feedback dataset.
ii. Language Detection: Comments were categorized by language using *langdetect*.
iii. Data Visualization: *matplotlib* and *seaborn* visualized frequency in word counts.

2. Removal of Blank and Noisy Comments

i. Dataset Loading: The raw dataset was loaded into a *pandas* DataFrame from an Excel file.
ii. Data Cleaning: Blank comments, comments with only symbols, and comments containing only numbers were filtered out to ensure high-quality data.
iii. Data Export: The cleaned dataset was saved and downloaded for subsequent step, such as language translation.

3. Translation to English

i. Dataset Upload: The cleaned dataset was uploaded into Google Colab for translation.
ii. Translation Process: Using googletrans, non-English text was translated to English, ensuring consistency across the dataset.
iii. Dataset Export: The translated data was saved and exported as a new Excel file for further analysis.

4. Overall Sentiments and Trends

i. Data Import: The customer feedback data, stored in an Excel file, which was pre-translated for uniformity was loaded into a Pandas DataFrame.
ii. Data Cleaning: A custom cleaning function was implemented to preprocess the feedback text. This function converted the text to lowercase, removed URLs, stripped non-alphabetic characters, and eliminated extraneous spaces. The cleaned feedback was stored in a new column of the DataFrame.
iii. Sentiment Analysis Using VADER: The sentiment of each piece of feedback was evaluated using the VADER sentiment analysis tool from the *nltk* library. Feedback was categorized into three sentiment classes: Positive, Negative, and Neutral, based on the compound sentiment score.
iv. Sentiment Trend Analysis: Sentiment trends over time were analyzed by extracting the month and year from the "Response Date" column and grouping the feedback by month. The sentiment counts were visualized through line charts showing the monthly distribution of Positive, Negative, and Neutral sentiments.
v. Visualization: Various visualizations were created to illustrate sentiment distribution and trends. A pie chart was generated to show the overall sentiment breakdown, while separate line charts were produced for positive, negative, and

neutral sentiments, respectively. These trends provided insights into shifts in customer perception over time.

5. Text Mining and Sentiment Analysis with Trends

i. Data Collection: The dataset containing customer feedback was uploaded into the Python environment (Google Colab) for further processing. This dataset included comments that were pre-translated for uniformity.
ii. Data Cleaning: The data underwent basic text preprocessing, which involved cleaning the comments by removing special characters, numbers, and extra spaces. All text was converted to lowercase for uniformity. Further, comments were lemmatized, and stop-words were removed for topic modeling purposes to ensure relevant words are captured.
iii. Topic Modeling (LDA): For topic extraction, the LDA method was applied using the cleaned comments. The number of topics was determined on coherence scores, which were computed for varying topic numbers to ensure the most meaningful and interpretable results. Utilizing word cloud, the topics were manually labeled and interpreted based on the top words identified by the LDA model. These topics were grouped into distinct themes related to the customer feedback.
iv. Sentiment Analysis: Sentiment analysis was performed using a pre-trained BERT model. Each comment was classified into sentiment categories ranging from strongly negative (1 star) to strongly positive (5 stars). The sentiment of each comment was then aggregated by topic, providing a sentiment distribution across different customer experience themes.
v. Aspect-Based Sentiment Analysis (ABSA): The sentiment for each topic was aggregated to understand how sentiment trends varied across different aspects of the feedback. A bar plot visualized the sentiment distribution for each topic.
vi. Trend Visualization: Temporal analysis was conducted by extracting monthly trends for both topic distribution and sentiment. These trends were visualized through line plots, showing how sentiment and topic frequencies evolved over time.
vii. Results and Visualization: The final dataset, which included topics, sentiments, and trends over time, was saved as a new file. Visualization results, such as word clouds for topics and sentiment trend plots, were generated to illustrate the insights derived from the analysis.

3.8 Reliability Check of Predicted Sentiment Labels

To assess the reliability of VADER and BERT's predicted sentiment labels, the following steps were undertaken:

1. A random sample of 500 feedback entries was selected from the dataset, ensuring an unbiased representation of the data.
2. Each feedback entry in the sample was manually labeled with its correct sentiment (positive, neutral, negative for VADER, and 1,2,3,4,5 for BERT) to serve as the ground truth.

3. The VADER-predicted labels for the same 500 feedback entries were compared against the manually assigned ground truth labels. Similarly, the BERT-predicted labels for 500 feedback entries were compared against the manually assigned ground truth labels.
4. Then, reliability metrics, including accuracy, precision, recall, and F1-score, were calculated in Google Colab to evaluate the model's performance (refer table 1 below).

Table 1: Reliability Metrics of Sentiment Labels

	Accuracy	*Precision*	*Recall*	*F1-Score*
VADER	82%	86%	82%	82%
BERT	86%	86%	86%	85%

Key Insights: VADER

- VADER shows high reliability in sentiment prediction for this dataset, with strong precision indicating confidence in the correctness of its predictions.
- The results suggest that VADER is effective at classifying sentiments that are clearly expressed.
- The recall indicates that the model might miss some nuanced or less obvious sentiments. This could stem from challenges like handling mixed sentiments, sarcasm, or domain-specific language.
- Misclassifications might occur in neutral feedback, which can often be more subjective or context-dependent.
- With an overall accuracy of 82%, VADER is a reliable tool for sentiment analysis in this context.

Key Insights: BERT

- With an accuracy of 86%, BERT demonstrates strong reliability in classifying sentiments. Its ability to achieve balanced precision and recall highlights that it is equally adept at identifying correct predictions and avoiding false ones.
- When compared to VADER (Accuracy: 82%, Precision: 86%, Recall: 82%, F1-Score: 82%), BERT shows a noticeable improvement in overall accuracy and recall. This suggests that BERT is better at handling nuanced or complex sentiments.
- The close alignment of precision (0.86) and recall (0.86) shows that the model has minimal trade-offs between identifying true positives and avoiding false positives. This is indicative of a well-calibrated model.
- While the metrics are high, the F1-Score (0.85) being slightly lower than precision and recall suggests minor misclassifications. These could stem from: ambiguity in middle-range sentiments (e.g., distinguishing between 3 stars and 4 stars); and feedback with mixed sentiments or subtle language variations.
- For applications like customer feedback analysis in the hospitality domain, BERT's reliability at 86% makes it a dependable tool. It provides actionable insights while minimizing errors in sentiment classification.

3.9 Ethical Considerations

Given that the study involves the analysis of internal customer feedback, ensuring ethical handling of the data is critical. The following steps will be taken to address ethical concerns:

- **Data Privacy**: The feedback data will be anonymized to protect the identity of guests. Any personally identifiable information (PII) will be removed during the data extraction process.
- **GDPR Compliance**: The study will comply with the General Data Protection Regulation (GDPR) guidelines, ensuring that guest data is handled in a lawful and transparent manner.
- **Anonymization**: As the data used in the study is internal, anonymization protocols will ensure that the feedback cannot be traced back to individual guests, safeguarding privacy and confidentiality (Prabhakar, 2023).

4. Data Analysis and Results

This chapter presents a comprehensive analysis of the internal customer feedback data using text mining, sentiment analysis, and trend analysis. The results provide a detailed view of customer perceptions, highlighting key insights for enhancing service quality and customer experience.

4.1 Statistical Analysis of the Dataset

Before exploring into sentiment and topic analysis, an initial statistical assessment was conducted to ensure data quality and understand the dataset's structure. This analysis provided context for the subsequent results, highlighting key characteristics such as word count distribution and language diversity.

4.1.1 Blank Comments Analysis

- Number of Blank Comments: 214
- Explanation: These comments were excluded from the analysis as they contained no substantive content. Removing blank entries ensures the dataset remains focused on meaningful feedback, enhancing the accuracy and reliability of the text mining and sentiment analysis processes.

4.1.2 Word Count Statistics

Understanding the distribution of word counts in customer feedback provides insights into the level of detail provided by respondents.

Table 2: Word Count Statistics

Statistic	Value
Total Non-Blank Comments	9726
Mean Word Count	26
Standard Deviation	28
Minimum Word Count	1
1st Quartile (25%)	8
Median (50%)	17
3rd Quartile (75%)	33
Maximum Word Count	376

Explanation (Refer Table 2 and Figure 1):

- The wide range of word counts (from 1 to 376 words) indicates variability in feedback depth.
- The median word count of 17 suggests that most comments are concise but informative, while the upper quartile contains more detailed reviews.

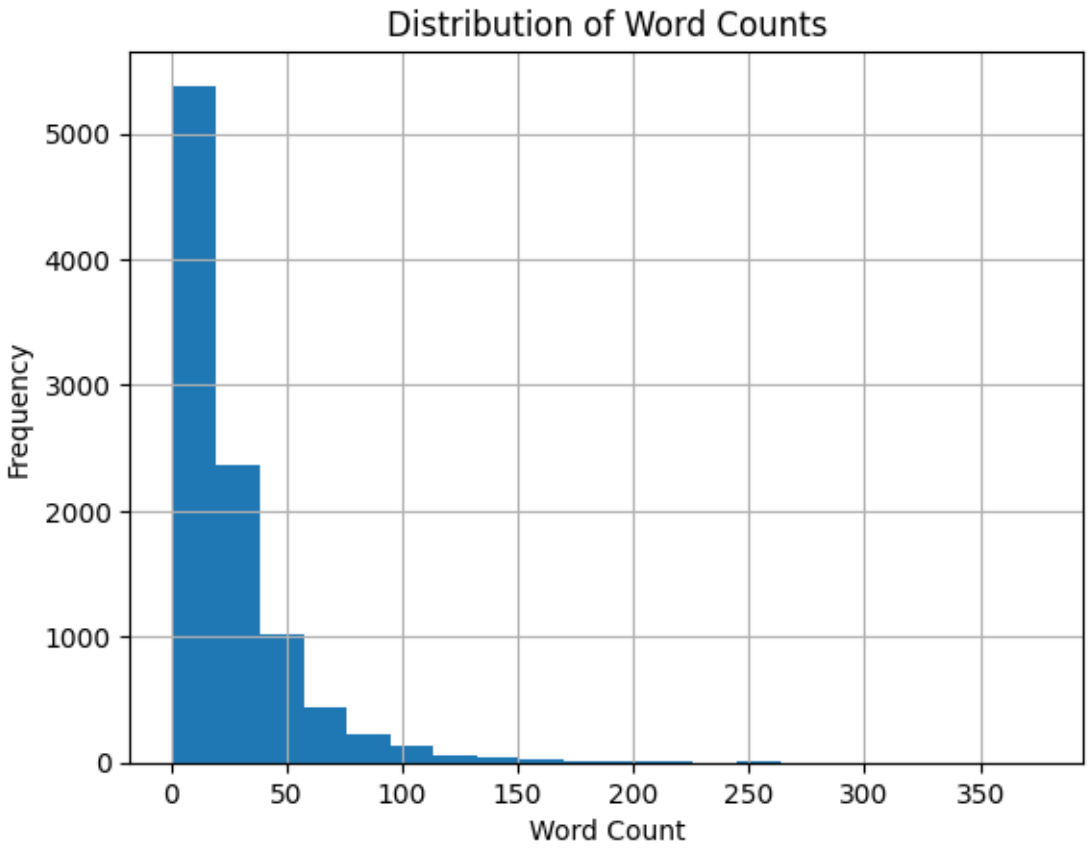

Figure 1: Word Count Distribution

4.1.3 Language Detection Results

The dataset's multilingual nature highlights the diversity of the customer base and the need for translation to ensure consistent analysis.

Table 3: Language Distribution of Customer Feedback

Language	Number of Comments
Finnish (fi)	7049
English (en)	1599
Swedish (sv)	291
Unknown	227
German (de)	152
Norwegian (no)	94
French (fr)	40
Danish (da)	40
Estonian (et)	38
Dutch (nl)	26
Afrikaans (af)	25
Somali (so)	24
Spanish (es)	22
Tagalog (tl)	18
Slovak (sk)	11
Swahili (sw)	11
Italian (it)	10
Portuguese (pt)	7
Catalan (ca)	6
Romanian (ro)	6
Welsh (cy)	6
Lithuanian (lt)	5
Vietnamese (vi)	5
Polish (pl)	4
Indonesian (id)	3
Japanese (ja)	2
Croatian (hr)	1
Albanian (sq)	1
Czech (cs)	1
Turkish (tr)	1
Hungarian (hu)	1

Explanation (Refer Table 3 and Figure 2):

- Dominant Language: Finnish, accounting for 72.5% of the feedback, indicating the primary customer demographic.
- Translation Requirement: Non-English comments were translated to English to ensure uniformity in sentiment and topic analysis.

- Removal of Unknown Language: Upon manual inspection, all the 227 unknown feedbacks were found noisy (i.e. containing symbols and punctuation marks), hence removed.

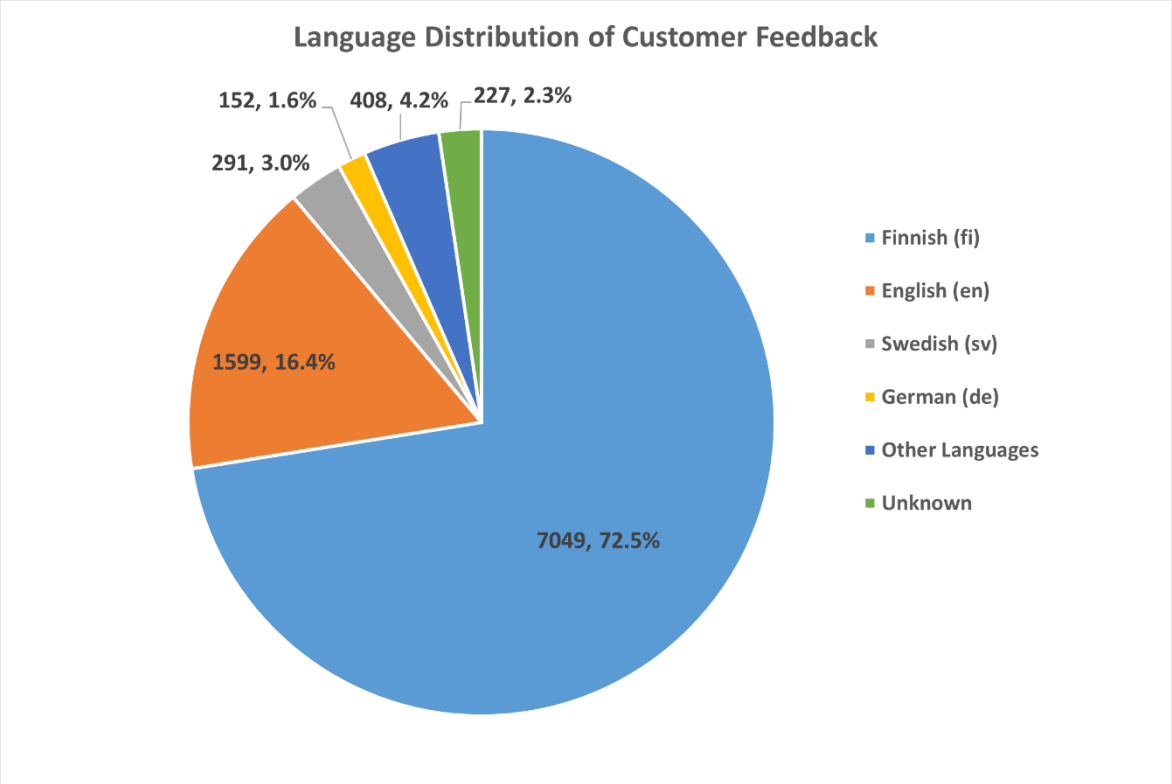

Figure 2: Language Distribution of Customer Feedback

4.2 Sentiment Analysis Results

4.2.1 Overall Sentiment Distribution

The sentiment analysis categorizes feedback into positive, negative, and neutral sentiments to understand customer perceptions.

Key Insights (Refer Figure 3):

- Positive Sentiment Dominance: **58.2%** of the feedback is positive, indicating that the majority of internal customers are satisfied or have a favorable view. This suggests overall positive experiences or satisfaction with the services.
- Negative Sentiment: **25.7%** of the feedback is negative. While this is a smaller proportion compared to positive feedback, it represents a significant segment that might highlight areas for improvement. Analyzing the specific concerns within this group could uncover critical issues needing attention.

- Neutral Sentiment: **16.1%** of the feedback is neutral. This indicates that a portion of the feedback does not convey strong positive or negative feelings. These comments might reflect mixed experiences or lack of strong opinion.

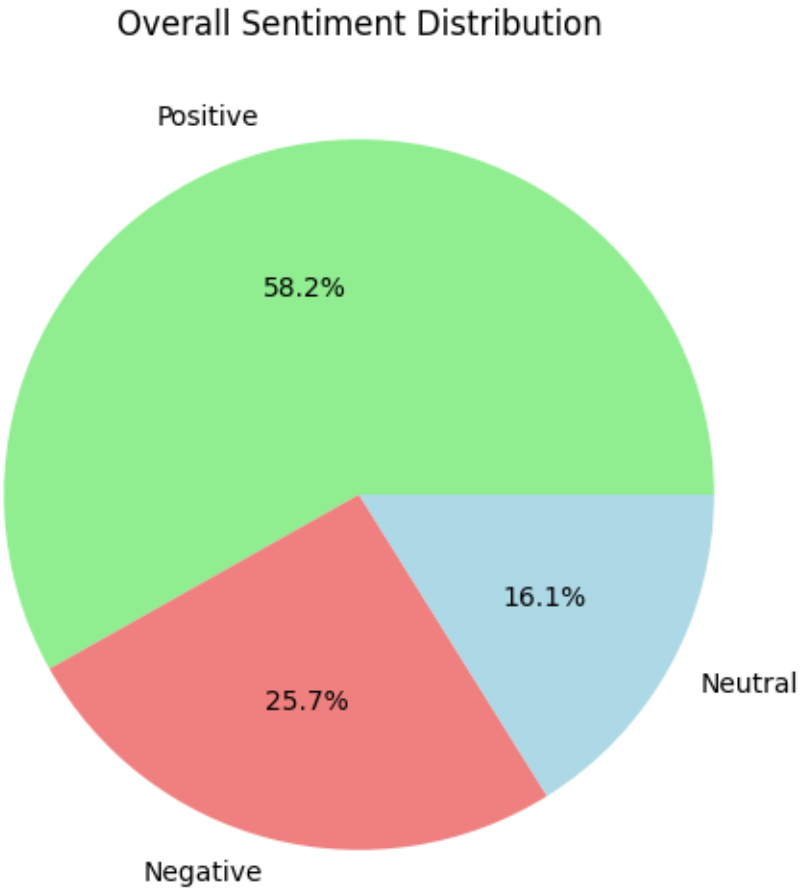

Figure 3: Overall Sentiment Distribution of Customer Feedback

4.2.2 Sentiment Trends Over Time

This section analyzes how sentiment trends evolve over the study period, highlighting seasonal or event-based variations.

Key Insights (Refer Figure 4):

- Positive Sentiment Dominance: Throughout the observed period (Nov 2021-Oct 2024), positive feedback consistently remains the highest across all months. Peaks in positive feedback are observed around the start of 2022 and mid-2023, suggesting periods of high customer satisfaction. The overall trend shows a slight decline towards the end of the period but stabilizes later.
- Fluctuations in Negative Feedback: Negative sentiment remains relatively low but shows noticeable fluctuations. There is a gradual increase in negative feedback around the middle of 2022 and early 2023, which may correspond to specific

operational challenges or service issues. A noticeable spike in negative feedback during some months aligns with a sharp decline in positive sentiment, suggesting possible service disruptions or customer dissatisfaction.

- Neutral Sentiment Stability: The neutral feedback trend is relatively stable with minor fluctuations. A slight increase in neutral feedback during mid-2022 and 2023 might indicate periods of mixed or indifferent customer experiences.
- Overall Trends: The trend suggests that positive feedback outweighs negative and neutral sentiments consistently, indicating that the hotel's service quality is generally well-received. However, the dips in positive sentiment, especially towards late 2023, may suggest areas where service improvements are necessary to maintain customer satisfaction.

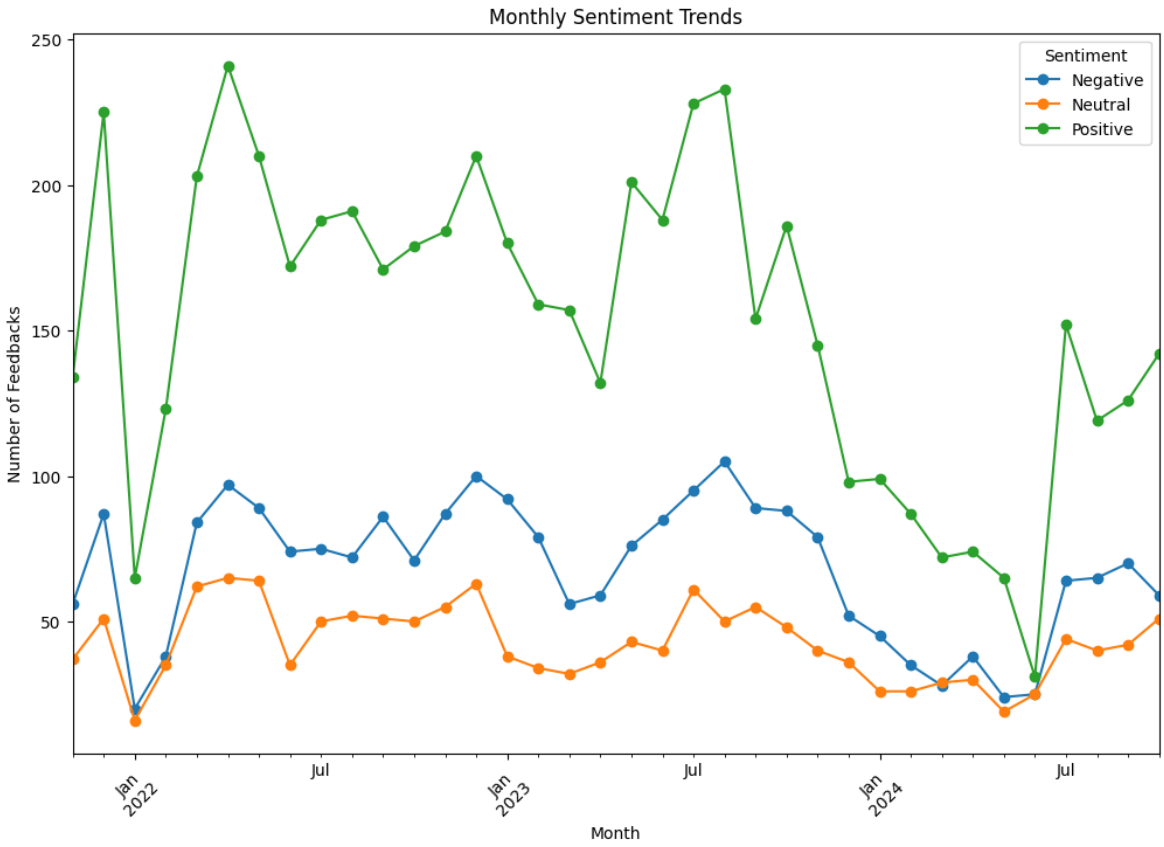

Figure 4: Monthly Sentiment Trends

4.3 Topic Modeling Results

4.3.1 Optimal Number of Topics

The Latent Dirichlet Allocation (LDA) model was used to extract key themes from customer feedback. The optimal number of topics (which is 8 in our case) was determined based on coherence scores (refer figure 5).

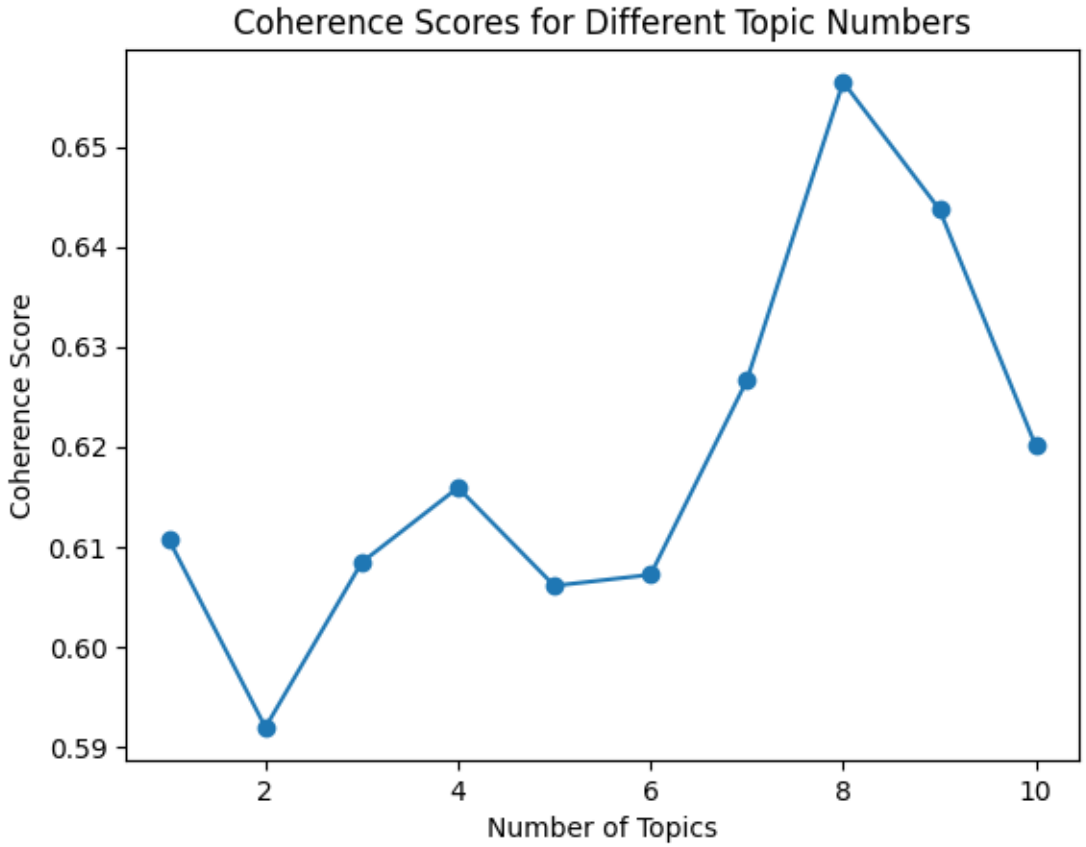

Figure 5: Coherence Score vs. Number of Topics

4.3.2 Identified Topics

The keywords obtained from topic modeling were manually interpreted using word clouds to identify key phrases for each topic.

Figure 6: Word Clouds for Topic 1, Topic 2, Topic 3 and Topic 4

Topic 1: Breakfast Experience

- Keywords (refer figure 6): breakfast, table, food, time, coffee, good, end, morning, dish, space

Topic 2: Room & Staff Service

- Keywords (refer figure 6): room, hotel, good, stay, staff, door, make, clean, leave, clothes

Topic 3: Check-in Experience

- Keywords (refer figure 6): time, room, check, long, queue, ask, respa, hotel, answer, book

Topic 4: Value for Money

- Keywords (refer figure 6): breakfast, hotel, good, nice, free, room, price, staff, hotel, time

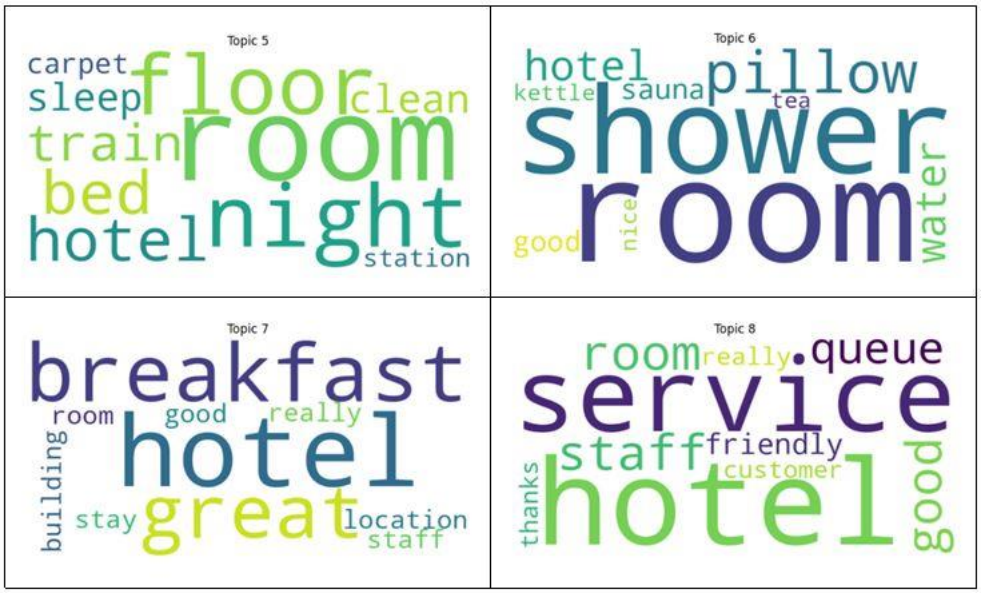

Figure 7: Word Clouds for Topic 5, Topic 6, Topic 7 and Topic 8

Topic 5: Room Comfort & Noise

- Keywords (refer figure 7): room, night, floor, hotel, bed, train, clean, sleep, carpet, station

Topic 6: Room Amenities

- Keywords (refer figure 7): room, shower, pillow, hotel, water, good, sauna, kettle, nice, tea

Topic 7: Overall Experience & Location

- Keywords (refer figure 7): hotel, breakfast, great, location, stay, room, good, staff, building, really

Topic 8: Customer Service Quality

- Keywords (refer figure 7): hotel, service, staff, good, room, queue, friendly, customer, really, thanks

4.4 Aspect-Based Sentiment Analysis (ABSA)

This analysis provides a granular view of customer sentiment on the above identified eight topics.

Key Insights (Refer Figure 8):

1. Breakfast Experience:
    - ✓ Balanced Feedback (3 Stars): The majority of comments fall under 3-star ratings, followed closely by 2-star ratings.
    - ✓ Negative Feedback: A notable portion of feedback is distributed across lower ratings (1 star), indicating some dissatisfaction with the breakfast experience, possibly related to variety, quality, or service issues.
2. Room & Staff Service:
    - ✓ Mixed Reviews: Most feedback is concentrated in the 3-star range, with significant 2-stars and 4-stars feedback.
    - ✓ Potential Issues: Lower ratings (1-2 stars) suggest areas where staff service or room quality could be improved, such as responsiveness or cleanliness.
3. Check-in Experience:
    - ✓ Balanced Distribution: Predominantly 3-star feedback, followed 2-stars and 1-star reviews.
    - ✓ Improvement Areas: The higher proportion of 2-star ratings suggests room for improvement in the check-in process, possibly related to efficiency or customer interaction.
4. Value for Money:
    - ✓ Mixed Sentiment: Distribution across 2-stars, 3-stars, and 4-stars ratings indicates differing opinions about whether customers felt they received good value.
    - ✓ Customer Expectation Management: Addressing value perception through service improvements or clearer pricing might enhance satisfaction.
5. Room Comfort and Noise:
    - ✓ Significant Dissatisfaction: A large proportion of feedback falls under 2-stars ratings, followed by 1-star, indicating widespread issues with comfort and noise levels.
    - ✓ Operational Focus: Reducing noise disturbances and improving room comfort should be prioritized to enhance customer satisfaction.
6. Room Amenities:
    - ✓ Moderate Satisfaction: Most feedback is concentrated in the 3-star range, with noticeable lower and higher ratings.

- ✓ Improvement Opportunity: Enhancing room amenities or addressing maintenance issues could improve customer perceptions.

7. Overall Experience & Location:
   - ✓ Highly Positive Feedback: This topic has a significantly higher proportion of 5-star ratings, suggesting strong satisfaction with the hotel's location and overall experience.
   - ✓ Strength Area: Highlighting location as a key marketing point while maintaining overall service quality will sustain positive sentiment.
8. Customer Service Quality:
   - ✓ High Satisfaction: 5-star and 4-star ratings dominate this topic, indicating strong customer service.
   - ✓ Maintaining Standards: Continue focusing on excellent service delivery while addressing occasional low ratings to further enhance overall customer experience.

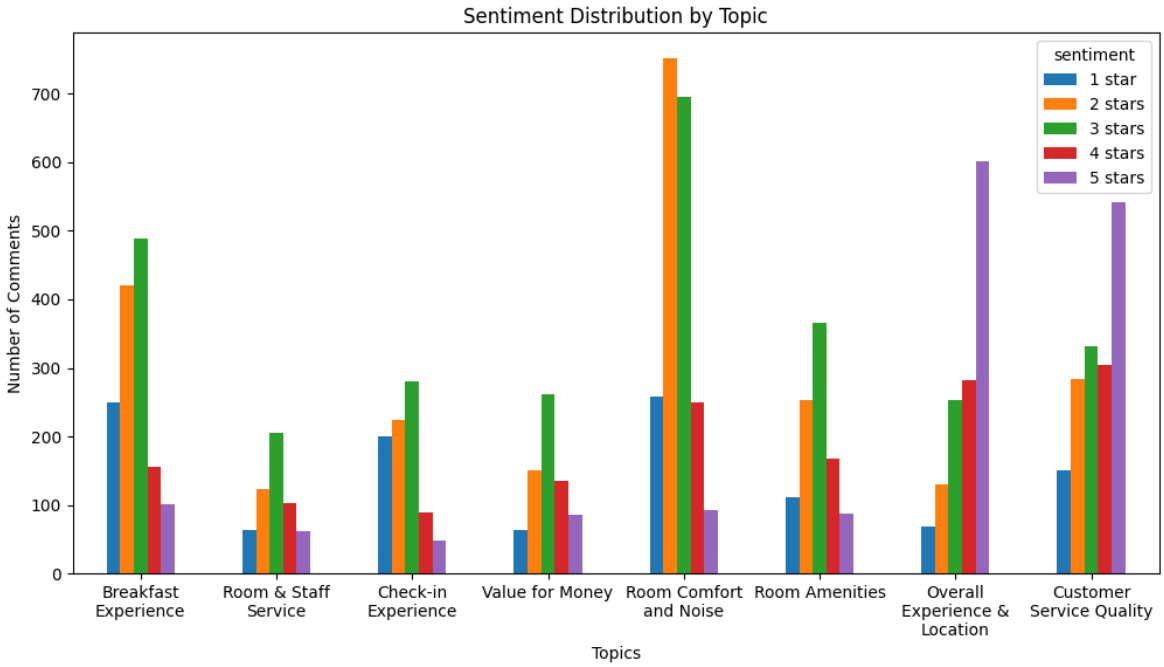

Figure 8: Sentiment Distribution by Aspect

4.5 Sentiment Trends Over Time for Identified Topics

This section analyzes the sentiment trends over time for each of the eight identified topics. Understanding how customer sentiment has evolved across various aspects of hotel operations provides valuable insights into areas of improvement and consistency. The trends are visualized through line charts.

1. Breakfast Experience (refer figure 9):

- Dominance of 3-Star and 2-Star Feedback: 3-star ratings remain consistently high, indicating that many customers view the breakfast experience as average. 2-star

ratings show regular peaks, suggesting recurring dissatisfaction with certain aspects of the breakfast service.

- Low Positive Feedback (4 and 5 Stars): 4-star and 5-star ratings are minimal throughout the timeline, with occasional increases. This suggests that fewer customers had an excellent or highly satisfying breakfast experience. Periodic peaks in 4-star ratings may indicate improvements or temporary enhancements in service or menu offerings.
- Negative Sentiment Trends: 1-star ratings show periodic spikes, suggesting some instances where the breakfast experience did not meet customer expectations. Peaks in 1-star and 2-star ratings often align, suggesting common issues related to service quality or food consistency.
- Volatility in Feedback Trends: The overall feedback trend demonstrates significant variability, especially in 2-star and 3-star ratings, indicating inconsistent breakfast service. Positive feedback remains relatively stable but low, suggesting room for improvement in maintaining consistently high standards.

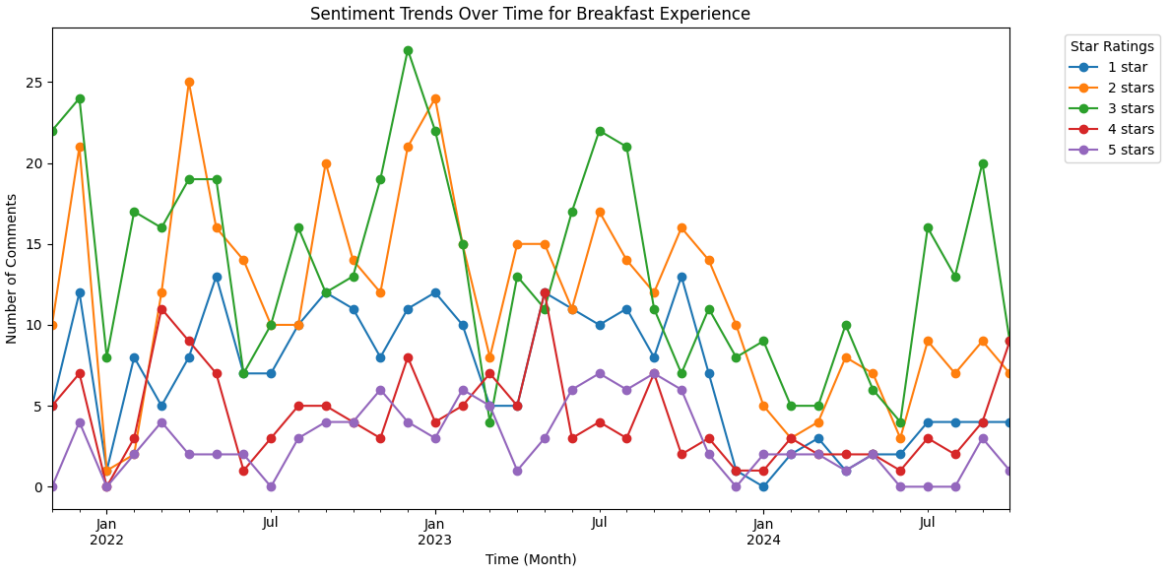

Figure 9: Sentiment Trends Over Time for "Breakfast Experience"

2. Room & Staff Service (refer figure 10):

- Consistent 3-Star Ratings: 3-star ratings are the most prevalent throughout the timeline, indicating a generally moderate or average perception of room and staff service. Peaks in 3-star feedback suggest that customers frequently identify areas that meet basic expectations but do not exceed them.
- Moderate Positive Sentiment (4 and 5 Stars): 4-star ratings show periodic peaks, suggesting some positive experiences, though not consistently high. 5-star ratings are minimal and sporadic, indicating that very few customers had exceptional experiences.
- Variable Negative Sentiment (1 and Stars): 2-star ratings show periodic increases, particularly during mid-2022 and late 2023, suggesting recurring issues in service quality or staff performance. 1-star ratings are less frequent but appear consistently, indicating isolated cases of dissatisfaction.

- Staff Service Improvement Areas: The overall variability in sentiment trends, especially in the 2-star and 3-star ratings, suggests that the quality of room and staff service is inconsistent. Positive feedback spikes (4 and 5 stars) indicate occasional service excellence, highlighting potential best practices that could be standardized.

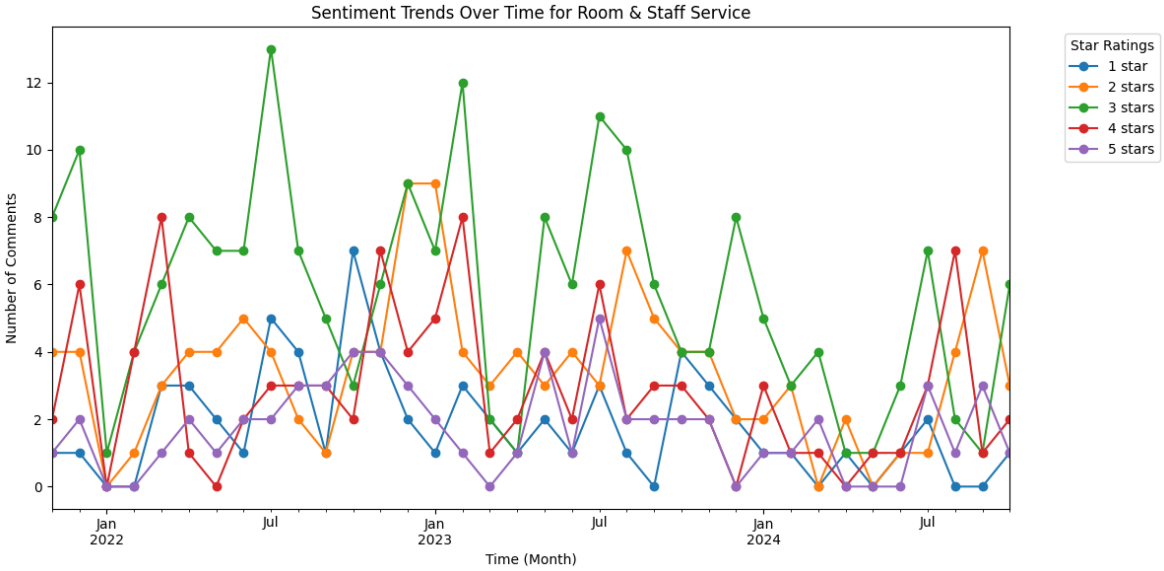

Figure 10: Sentiment Trends Over Time for "Room & Staff Service"

3. Check-in Experiences (refer figure 11):

- High Variability in Sentiments: There are significant fluctuations in comment counts across all ratings, indicating that customer experiences vary greatly over time. 3-star ratings dominate most months, implying that many customers feel their experiences are average or mixed.
- Negative Sentiment Spikes: 1-star and 2-star ratings consistently show peaks, with notable surges in early 2022 and mid-2023. This pattern suggests periodic dissatisfaction, possibly due to recurring operational challenges.
- Lower Positive Feedback: 4-star and 5-star ratings are consistently low compared to other ratings, suggesting fewer outstanding customer experiences during check-ins.
- Seasonal Patterns: Spikes in comments (both positive and negative) often occur around specific months, hinting at potential seasonal factors influencing sentiment.

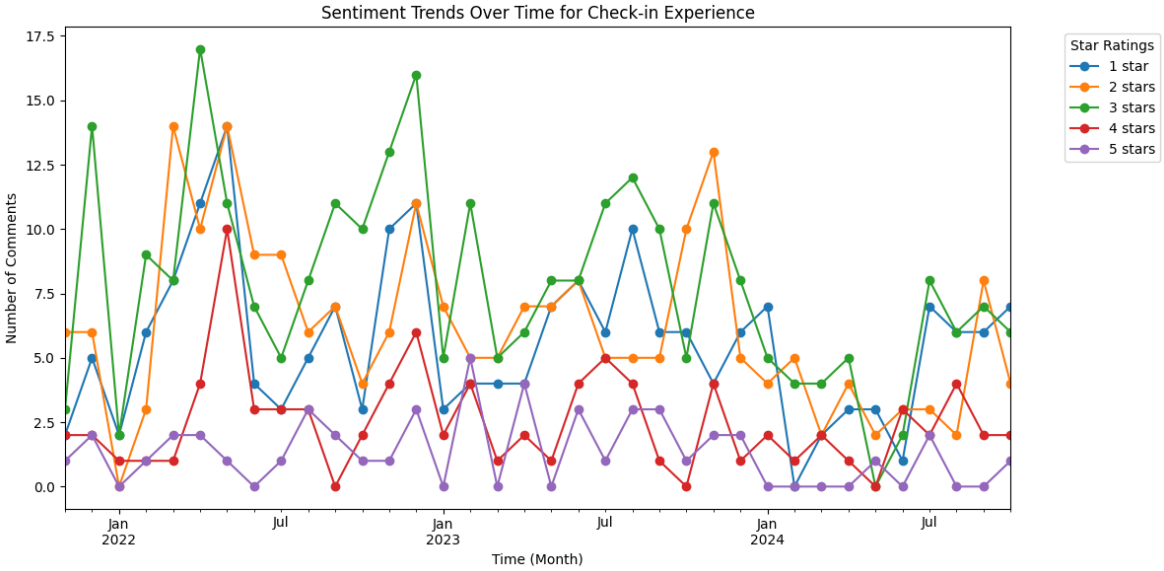

Figure 11: Sentiment Trends Over Time for "Check-in Experience"

4. Value For Money (refer figure 12):

- Dominance of Neutral Sentiments: 3-star ratings consistently exhibit the highest frequency, indicating customers often perceive the value for money as average or moderate.
- Recurring Negative Feedback: 1-star and 2-star ratings demonstrate regular spikes, especially during early 2022 and mid-2023, highlighting dissatisfaction during these periods. This might be linked to pricing changes or unmet expectations.
- Limited Positive Feedback: 4-star and 5-star ratings are consistently low, with only a few noticeable peaks, reflecting fewer customers who perceive the value as exceptional.
- Seasonal Trends: Periods of heightened feedback (both positive and negative) tend to align with specific times of the year, indicating that external factors (e.g., promotions or economic shifts) may influence sentiments about value for money.

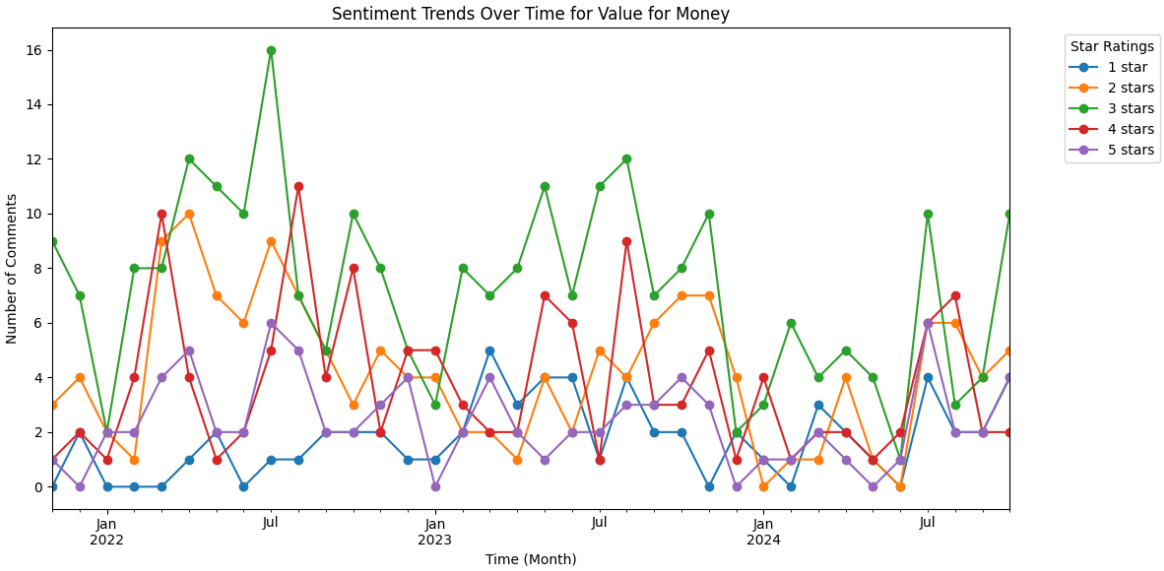

Figure 12: Sentiment Trends Over Time for "Value for Money"

5. Room Comfort & Noise (refer figure 13):

- Prevalence of Negative Feedback: 2-star ratings often surpass other categories, suggesting that a significant portion of customers perceive comfort and noise issues as below expectations. 1-star ratings also show periodic spikes, particularly in early 2022 and late 2023, reflecting heightened dissatisfaction during those times.
- Mixed Sentiments with 3-Star Ratings: 3-star ratings frequently trend high, indicating that many customers have neutral or average experiences related to room comfort and noise.
- Limited Positive Sentiments: 4-star and 5-star ratings are consistently low, showing few customers express high satisfaction with comfort and noise.
- Seasonal and Recurring Patterns: Spikes in comments, both negative (1-star and 2-star) and neutral (3-star), often occur at similar times each year, suggesting that seasonal factors (e.g., high occupancy periods) may contribute to comfort and noise issues.

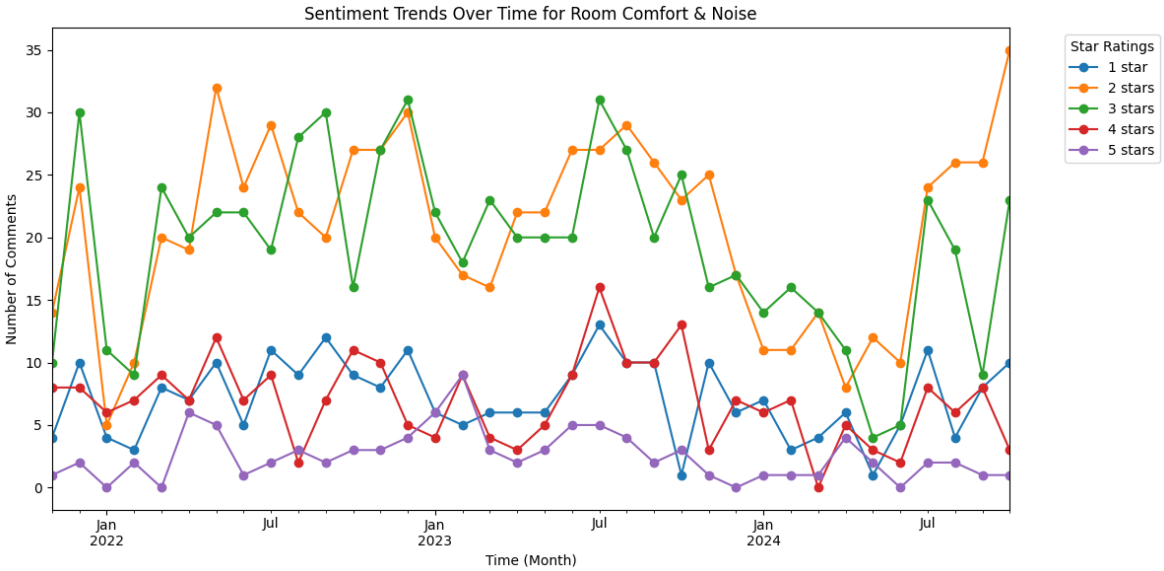

Figure 13: Sentiment Trends Over Time for "Room Comfort & Noise"

6. Room Amenities (refer figure 14):

- 3-Star Ratings (green line): The 3-star ratings consistently have the highest number of comments, with noticeable peaks around mid-2022 and late 2023. This indicates that most guests are moderately satisfied with room amenities and frequently provide feedback.
- 2-Star Ratings (orange line): The 2-star ratings also show significant fluctuations, with a peak around late 2023. This suggests periods where guests found room amenities below their expectations.
- 1-Star Ratings (blue line): The 1-star ratings have a relatively lower and more stable number of comments, with occasional peaks. This indicates a smaller number of guests who were very dissatisfied with room amenities.

- 4-Star Ratings (red line): The 4-star ratings show moderate variability with fewer comments compared to 3-star and 2-star ratings, indicating good satisfaction levels, but less frequent feedback.
- 5-Star Ratings (purple line): The 5-star ratings have the lowest number of comments overall, with minimal variability and fewer peaks. This suggests that fewer guests rated room amenities as excellent.
- Overall Trend: There is a general trend of increasing comments towards the end of the period for most star ratings, especially for 3-star and 2-star amenities. This could reflect growing engagement or changes in guest experiences over time.

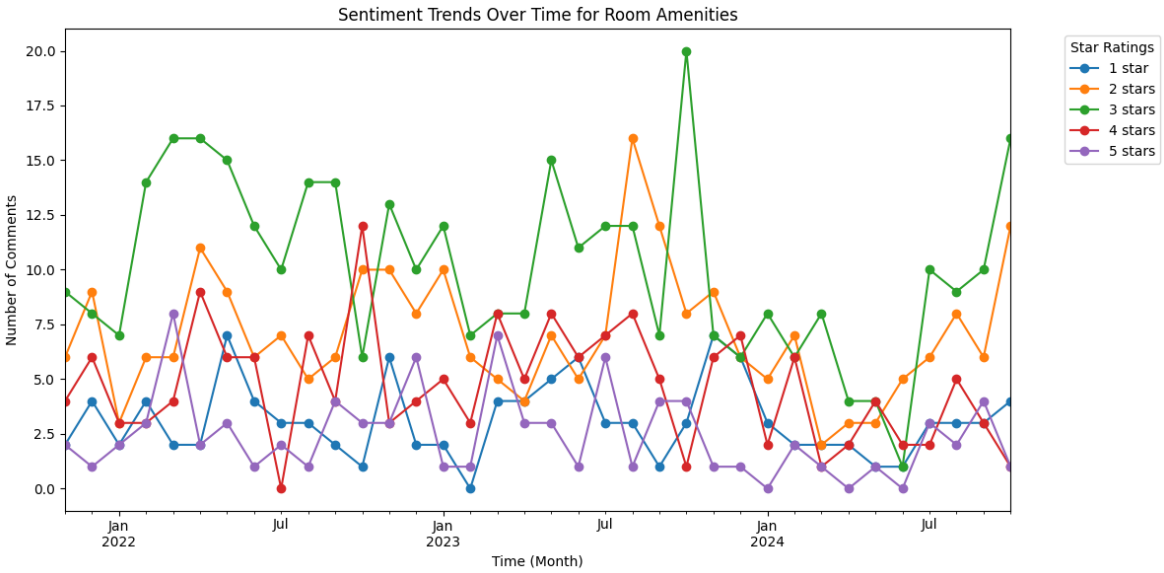

Figure 14: Sentiment Trends Over Time for "Room Amenities"

7. Overall Experience & Location (refer figure 15):

- 5-Star Ratings: The purple line consistently shows the highest number of comments throughout the period, indicating that most guests rated their overall experience and the hotel's location very positively. Peaks in the graph, reaching up to 26 comments, suggest particularly high satisfaction during certain months.
- 4-Star Ratings: The red line shows significant fluctuations, with notable peaks reaching around 17 comments. This indicates that while many guests were also highly satisfied, there were more varied experiences compared to the 5-star ratings.
- 3-Star Ratings: The green line represents moderate and relatively stable trends, with the number of comments generally staying below 10. This suggests an average level of satisfaction among some guests, indicating consistent but not outstanding experiences.
- 2-Star Ratings: The orange line also shows a stable trend with relatively fewer comments, staying below 10. This indicates that some guests had below-average experiences with the overall experience and location.

- 1-Star Rating: The blue line shows the lowest number of comments, rarely exceeding 5. This suggests that only a small portion of guests had poor experiences regarding the hotel's overall experience and location.

Overall, the trend shows that the majority of guests were highly satisfied with the hotel's overall experience and location, as reflected by the dominance of 4-star and 5-star ratings. However, the presence of lower-star ratings highlights areas where improvements could be made to enhance guest satisfaction further.

Figure 15: Sentiment Trends Over Time for "Overall Experience & Location"

8. Customer Service Quality (refer figure 16):

- 5-Star Ratings: The number of 5-star comments is generally higher than other ratings, with noticeable peaks around mid-2022 and mid-2023, reaching above 25 comments. This indicates consistent high satisfaction with customer service quality during these periods.
- 1-Star Ratings: The number of 1-star comments is relatively low and stable, with occasional peaks but generally staying below 10 comments. This suggests that while some guests were dissatisfied, it was a minority.
- 2-Star Ratings: The 2-star comments show fluctuations but remain mostly below 15 comments, with some peaks around mid-2022 and mid-2023. This indicates occasional dissatisfaction, but not as widespread as higher ratings.
- 3-Star Ratings: The 3-star comments also fluctuate, with peaks around mid-2022 and mid-2023, but generally stay below 15 comments. This suggests moderate satisfaction among some guests, with variability over time.
- 4-Star Ratings: The 4-star comments show similar trends to the 3-star comments, with peaks around mid-2022 and mid-2023, staying mostly below 15 comments. This indicates a good level of satisfaction, just below the highest rating.

The graph indicates that 5-star ratings are the most frequent, highlighting strong customer satisfaction with service quality. There are noticeable peaks in comments for all ratings around mid-2022 and mid-2023, which may correspond to specific events or

changes in service during those periods. The lower frequency of 1-star ratings suggests that most guests had positive experiences with customer service.

Figure 16: Sentiment Trends Over Time for "Customer Service Quality"

4.6 Temporal Topic Trends of Customer Feedback

This subchapter examines the trends in guest comments across the eight identified topics over time, as visualized in Figure 23. The analysis provides insights into shifting guest priorities and areas of concern, which are critical for understanding customer experiences and refining service quality in the hotel.

Overview of Trends

The line chart in Figure 17 depicts the monthly distribution of guest comments for the topics: Breakfast Experience, Room & Staff Service, Check-in Experience, Value for Money, Room Comfort & Noise, Room Amenities, Overall Experience & Location, and Customer Service Quality. Each topic displays unique temporal patterns, reflecting varying levels of guest focus and engagement.

Key Observations

1. Breakfast Experience and Customer Service Quality Lead Commentary
   Topics such as Customer Service Quality (grey line) and Breakfast Experience (blue line) consistently attract the highest number of comments across the analyzed period. Peaks in Breakfast Experience are particularly noticeable during late summer and early winter months, potentially coinciding with peak tourist seasons when breakfast becomes a critical part of the guest experience.
2. Fluctuations in Check-in Experience and Room Comfort & Noise
   The volume of comments related to Check-in Experience (green line) and Room Comfort & Noise (purple line) show significant variability. For instance, notable spikes in Check-in Experience occur intermittently, possibly aligning with

operational changes or events that impacted the check-in process. Similarly, Room Comfort & Noise displays sharp peaks during colder months, suggesting seasonal sensitivity to room conditions.

3. Steady Engagement with Value for Money and Room Amenities
   While Value for Money (red line) and Room Amenities (brown line) maintain relatively lower levels of commentary, their patterns remain steady over time, with occasional surges. This consistency highlights these topics as underlying concerns but not dominant focal points for most guests.
4. Overall Experience & Location Sees Seasonal Variations
   Comments regarding Overall Experience & Location (pink line) exhibit a seasonal pattern, with a notable rise during the summer months. This trend may reflect the influx of leisure travelers who prioritize location for sightseeing and overall trip satisfaction during the warmer seasons.
5. Event-Specific Spikes and Drops
   A distinct dip across almost all topics is evident in early 2024, followed by sharp rebounds. This could correspond to a temporary operational adjustment or an external factor affecting guest commentary, such as a period of reduced occupancy or renovations.

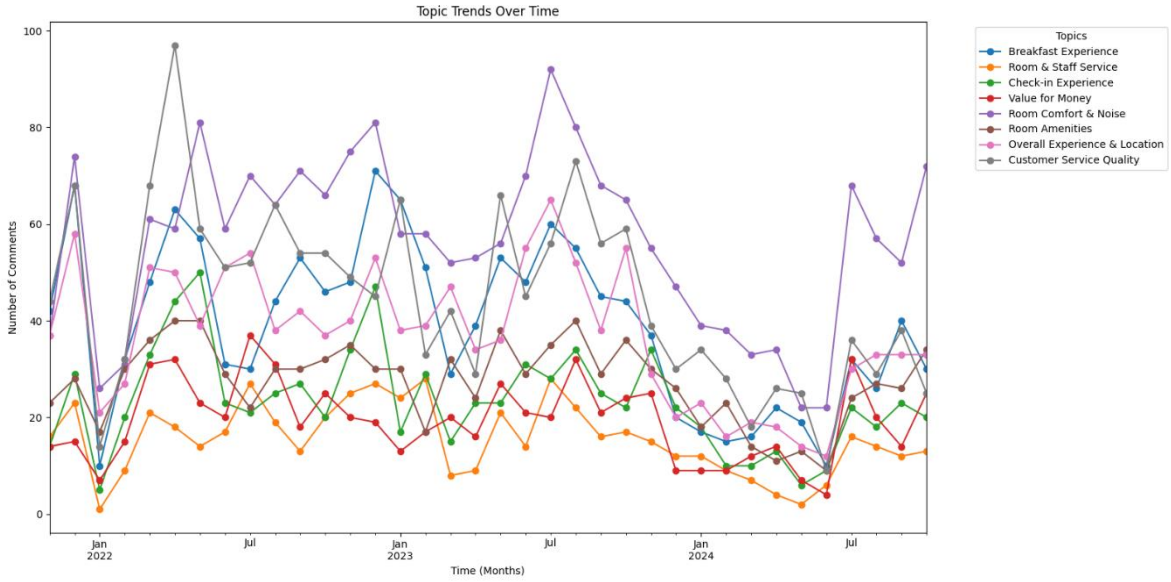

Figure 17: Topic Trends Over Time

5. Findings and Discussion

This chapter synthesizes the insights derived from the data analysis and results in Chapter 4, contextualizing them within the objectives of the study and the broader hospitality literature. The findings reflect how text mining and sentiment analysis provided actionable

insights into customer perceptions, and how these insights align with or diverge from existing research.

5.1 Summary of Findings

The analysis of customer feedback revealed several notable trends and patterns across the eight identified topics. These findings shed light on customer priorities, seasonal variations in sentiment, and recurring concerns that significantly influence the guest experience. Key results include:

1. **Breakfast Experience**: Frequent commentary with a mix of sentiments, highlighting variability in guest satisfaction.
2. **Room & Staff Service**: Mixed feedback, with significant positive sentiments tempered by specific negative experiences.
3. **Check-in Experience**: Mixed feedback, indicating potential inefficiencies or inconsistencies in the check-in process.
4. **Value for Money**: Divergent opinions, with some customers appreciating the value and others expressing dissatisfaction.
5. **Room Comfort & Noise**: Persistent dissatisfaction, with significant negative feedback regarding noise and comfort levels.
6. **Room Amenities**: Moderate satisfaction levels, with variability in feedback tied to specific amenities such as water, tea, and sauna facilities.
7. **Overall Experience & Location**: Generally positive feedback, underscoring the importance of the hotel's location as a key strength.
8. **Customer Service Quality**: A strong area for the hotel, consistently garnering high positive feedback and praise.

5.2 Discussion of Key Trends

5.2.1 Breakfast Experience

Guest feedback on the breakfast experience revealed seasonal peaks in commentary, with a noticeable increase during tourist-heavy months. Common themes included the quality, variety, and presentation of food. While many guests expressed satisfaction, there was notable negative feedback about overcrowding and the availability of specific items during peak times. This aligns with literature emphasizing the role of food quality in shaping guest satisfaction (Bichler, Pikkemaat, & Peters, 2021).

5.2.2 Room & Staff Service

Guest opinions on room and staff service varied, with many appreciating the cleanliness and attentiveness of staff. Negative feedback often focused on delays in housekeeping and inconsistencies in staff behavior. These findings suggest the need for continuous monitoring and refinement of service standards.

5.2.3 Check-in Experience

Feedback on the check-in process reflected variability in guest experiences. While many guests found the process smooth and efficient, others expressed frustration with long queues and staff unpreparedness during busy periods. Such findings align with studies that identify check-in efficiency as a critical determinant of first impressions in hotel stays (Shao, Cho, Tse, & Zou, 2017).

5.2.4 Value for Money

Opinions on value for money varied widely, with sentiments often tied to the perceived quality of room amenities and services relative to pricing. Negative feedback occasionally highlighted discrepancies between expectations set by online promotions and the actual experience. These insights suggest a need for transparent pricing and consistent service delivery to enhance value perception.

5.2.5 Room Comfort & Noise

This category received some of the most negative feedback, particularly during colder months. Guests frequently mentioned disturbances from external noise and discomfort related to room temperature and furniture. These findings echo prior studies highlighting the impact of external noise on customer satisfaction in the hospitality industry (Oberson, 2023).

5.2.6 Room Amenities

Feedback on room amenities was mixed, with some guests praising the availability of conveniences like kettles and saunas, while others pointed out issues with maintenance and replenishment of supplies. This variability indicates an opportunity for operational improvements to enhance guest satisfaction consistently.

5.2.7 Overall Experience & Location

Overall experience and location were consistent strengths, with many guests highlighting the hotel's proximity to key attractions and transport hubs. Positive feedback dominated this category, reflecting the critical role of location in guest satisfaction (Yang, Mao, & Tang, 2017).

5.2.8 Customer Service Quality

Customer service emerged as a strong area for the hotel, with consistent positive sentiment. Guests praised staff for their friendliness, responsiveness, and professionalism. However, occasional negative feedback pointed to isolated incidents of miscommunication or delays in service delivery. These findings reinforce the importance of staff training and customer-centric service, as emphasized in the hospitality literature (Gerber, 2023).

5.3 Comparing Findings with the Literature

The findings align with existing research emphasizing the importance of service quality, comfort, and location in shaping guest satisfaction. Notably:

- **Breakfast Experience**: The variability in guest satisfaction mirrors prior studies that link meal quality to perceived value (Yrjölä, Rintamäki, Saarijärvi, Joensuu, & Kulkarni, 2019).
- **Customer Service**: Strong performance in this area corroborates findings that professional, friendly staff are a cornerstone of positive guest experiences (Chen, Chen, Liu, & Sharma, 2020).
- **Room Comfort & Noise**: Persistent issues in this area resonate with literature highlighting environmental factors as critical to satisfaction (Oberson, 2023).

However, the study also contributes novel insights, particularly regarding the application of text mining and sentiment analysis to uncover nuanced feedback trends over time.

6. Conclusions

This chapter provides a concise synthesis of the research, revisiting the objectives and summarizing key findings. It also highlights the contribution of the study to academic knowledge and practical applications in the hospitality industry.

6.1 Restating Research Objectives

The primary objective of this study was to explore how text mining and sentiment analysis can be utilized to enhance customer experience and service quality in hotel operations. The study aimed to achieve three specific objectives:

1. Determine the overall sentiment of customer feedback and analyze trends over time.
2. Identify the main topics emerging from customer feedback and evaluate their influence on guest sentiment and service quality.
3. Provide actionable recommendations based on findings to improve customer experience and operational efficiency.

By analyzing guest feedback from a Finnish hotel, the study sought to provide data-driven insights to improve decision-making in hospitality management.

6.2 Key Findings

The study successfully met its objectives through a comprehensive analysis of guest feedback. Key findings include:

1. **Overall Sentiment Analysis**:
    - A majority (58.2%) of feedback was positive, indicating overall guest satisfaction.
    - Negative feedback (25.7%) highlighted specific areas of concern, such as room comfort and noise, breakfast experience, and check-in efficiency.
    - Neutral feedback (16.1%) reflected mixed opinions or low engagement with specific aspects of the service.
2. **Topic Trends and Sentiment**:
    - **Customer Service Quality** was the most consistently positive topic, reinforcing its central role in guest satisfaction.
    - **Room Comfort & Noise** emerged as the most problematic aspect, with recurring negative feedback linked to environmental and operational factors.
    - **Breakfast Experience** and **Check-in Process** showed variability in guest satisfaction, influenced by seasonal peaks and operational adjustments.
    - The **Overall Experience & Location** topic consistently received high positive feedback, underscoring the strategic advantage of the hotel's location.
3. **Temporal Trends**:
    - Seasonal variations significantly influenced sentiment, with peaks in positive feedback during summer months and dips during winter.

- Negative feedback often aligned with operational challenges during peak tourist seasons or specific service disruptions.

4. **Novel Insights from Text Mining**:
    - The application of Latent Dirichlet Allocation (LDA) and sentiment analysis revealed granular insights into guest perceptions, allowing for the identification of actionable themes.

6.3 Contributions to Knowledge

This study makes several contributions to the fields of hospitality management and computational analysis:

1. Academic Contributions:
    - Demonstrates the utility of text mining and sentiment analysis in extracting actionable insights from unstructured guest feedback.
    - Highlights the importance of temporal analysis in understanding seasonal and operational influences on guest sentiment.
2. Practical Contributions:
    - Provides a framework for systematically analyzing guest feedback to identify strengths and weaknesses in hotel operations.
    - Offers actionable recommendations for improving customer experience, with a focus on addressing recurring issues and enhancing service delivery.

6.4 Implications for the Hospitality Industry

The findings underscore the critical role of leveraging technology to analyze customer feedback in the hospitality industry. Insights from this study can help hotels to:

- enhance customer experience by using sentiment analysis to proactively address guest concerns and improve service quality in high-impact areas such as breakfast services and room comfort.
- optimize operations by adjusting staff and resource allocation based on temporal trends and feedback patterns to improve operational efficiency.
- strengthen competitive advantage by highlighting strengths such as location and customer service in marketing strategies while addressing specific weaknesses to maintain a competitive edge.

6.5 Final Thoughts

This research demonstrated the potential of combining text mining and sentiment analysis to derive actionable insights from guest feedback, ultimately contributing to the enhancement of service quality and customer experience in the hospitality industry. By adopting data-driven strategies and addressing key areas of concern, the case-study hotel and similar establishments can foster greater customer satisfaction and loyalty in an increasingly competitive market.

7. Recommendations

This chapter outlines actionable recommendations for the case-study hotel based on the findings from this research. These recommendations focus on improving customer experience and operational efficiency while addressing specific areas of concern highlighted by the text mining and sentiment analysis.

7.1 Enhancing Hotel Operations

Improving Breakfast Experience

Menu Variety and Quality:

- Introduce a rotational menu with diverse options catering to various dietary preferences (e.g., vegetarian, vegan, gluten-free).
- Regularly update food items based on customer feedback trends to ensure freshness and relevance.

Service Efficiency:

- Allocate additional staff during peak hours to manage overcrowding and replenish food items promptly.
- Introduce pre-booking options for breakfast slots during high-demand periods to reduce wait times and improve flow.

Addressing Room Comfort and Noise

Noise Mitigation:

- Install soundproofing materials, such as acoustic panels, in rooms adjacent to noisy areas (e.g., near elevators or streets).
- Implement quiet hours and communicate noise policies to guests during check-in.

Comfort Enhancements:

- Regularly assess the quality of mattresses, pillows, and furniture for wear and tear, replacing them as needed.
- Offer temperature control options in rooms, such as portable heaters or air conditioning units, to address seasonal discomfort.

Refining Check-in Processes

Streamlining Workflow:

- Introduce self-check-in kiosks to expedite the process during busy periods.
- Enhance staff training to ensure they can handle peak check-in times effectively and resolve issues promptly.

Communication Improvements:

- Send automated reminders to guests with check-in instructions and estimated wait times before arrival.

Strengthening Value for Money

Transparency in Pricing:

- Ensure that online promotions and room descriptions accurately reflect what guests will receive to manage expectations.

Added Perks:

- Provide complimentary amenities, such as welcome drinks or free Wi-Fi upgrades, to enhance the perceived value of stays.

7.2 Enhancing Customer Experience

Elevating Customer Service Quality

Staff Training Programs:

- Conduct regular training sessions to emphasize empathy, active listening, and problem-solving skills among staff.
- Create scenario-based workshops to prepare staff for handling challenging customer interactions.

Feedback Utilization:

- Implement a real-time feedback system where guests can rate their experiences immediately, allowing the staff to address concerns on the spot.

- The hotel should aim to acknowledge and respond to every piece of customer feedback, whether positive, negative, or neutral. This practice ensures guests feel heard and valued, which can significantly enhance their perception of the hotel.

Utilizing Location Advantages

Local Partnerships:

- Collaborate with nearby attractions, restaurants, and transport services to offer exclusive discounts or packages to guests.

Promoting Accessibility:

- Clearly highlight transport options and nearby landmarks in marketing materials to attract location-conscious travelers.

7.3 Operational Recommendations

Monitoring and Managing Seasonal Variations:

- Use temporal analysis insights to adjust staffing, inventory, and services during peak tourist seasons.
- Plan for preventive maintenance during off-peak periods to minimize disruptions when demand is high.

Utilizing Technology for Continuous Improvement:

- Implement a data dashboard that visualizes ongoing feedback trends, enabling management to track and address issues in real time.

7.4 Economic and Implementation Roadmap: Cost and Complexity Alignment

To assist the case-study hotel in effectively implementing the recommendations, this section presents a roadmap that integrates cost considerations (low-cost, medium-cost, high-cost) with implementation complexity (easy, medium, hard). The recommendations are organized into short-term, mid-term, and long-term actions, providing a phased approach for addressing both immediate and strategic goals.

Short-Term Actions (0-6 Months)

Short-term actions focus on low-cost and easy-to-implement recommendations that can deliver immediate benefits with minimal financial or operational impact.

Low-Cost and Easy Recommendations:

- Improve the breakfast experience by using customer feedback to adjust menu options.

- Establish quiet hours to mitigate noise, communicated via in-room notices or check-in materials.
- Conduct in-house training for staff on customer service skills, emphasizing empathy and problem-solving.
- Use existing email or SMS systems to send automated reminders with check-in instructions and estimated wait times.

Mid-Term Goals (6-12 Months)

Mid-term actions involve medium-cost and moderately complex measures that require some resource investment but deliver significant improvements in operations and guest satisfaction.

Medium-Cost and Medium-Complexity Recommendations:

- Introduce basic self-check-in kiosks to streamline the check-in process.
- Implement a real-time feedback system to address guest concerns during their stay.
- Establish partnerships with local attractions, restaurants, and transport services to offer exclusive discounts or packages.
- Replace aging mattresses, pillows, and furniture as part of scheduled maintenance to enhance room comfort.

Long-Term Strategies (12+ Months)

Long-term actions encompass high-cost and complex initiatives that require careful planning and resource allocation. These measures are designed to provide sustainable, long-term benefits to the hotel's operations and guest experiences.

High-Cost and Hard Recommendations:

- Install soundproofing materials in rooms near noisy areas, such as streets or elevators.
- Upgrade HVAC systems for centralized climate control, ensuring consistent guest comfort.
- Expand or redesign the breakfast area to address overcrowding during peak hours.
- Invest in advanced sentiment analysis tools and dashboards for real-time tracking and continuous improvement based on guest feedback trends.

Across all phases, it is essential to monitor the outcomes of implemented actions and refine them based on guest feedback and operational data. This iterative approach ensures that investments—whether low-cost or high-cost—achieve their intended goals, and allows for adjustments to align with evolving priorities and constraints.

By implementing these recommendations, the case-study hotel can address key areas of guest dissatisfaction, build on its strengths, and position itself as a leader in customer experience within the hospitality industry. These actionable insights not only cater to the specific challenges faced by the hotel but also offer a replicable framework for improving service quality in similar establishments.-

7.5 Recommendations for Future Research and Development

While the study focuses on operational improvements, it also suggests potential areas for future exploration to further enhance service quality:

While this study provided valuable insights, it also opens avenues for further exploration:

1. Broader Datasets: Future research could expand the scope to include feedback from multiple hotels or diverse geographical locations for comparative analysis.
2. Advanced Sentiment Techniques: Incorporating more advanced deep learning models or multimodal analysis (e.g., combining text with images or audio) could yield more nuanced insights.
3. Real-Time Analysis: Investigating the potential of real-time sentiment analysis for operational decision-making could enhance responsiveness in hotel management.
4. Customer Segmentation: Segmentation of customer based on feedback patterns to understand different customer groups and their specific needs, allowing for more personalized and effective marketing and service strategies.
5. Predictive Analytics: Prediction of future customer satisfaction levels based on historical feedback data using machine learning models like regression analysis, decision trees, or neural networks to improve service based on predictive insights.
6. Comparative Analysis by Language: Comparison of feedback trends and sentiments across different languages to identify cultural or regional differences in customer expectations and satisfaction.

8. References

Adikari, A., Nguyen, S., Nawaratne, R., De Silva, D., & Alahakoon, D. (2024). Transforming customer digital footprints into decision enablers in hospitality. *Applied Sciences, 14*(7), 3114. https://doi.org/10.3390/app14073114

Bi, J.-W., Zhu, X.-E., & Han, T.-Y. (2024). Text analysis in tourism and hospitality: A comprehensive review. *Journal of Travel Research, 63*(1), 1-23. https://doi.org/10.1177/00472875241247318

Bichler, B.F., Pikkemaat, B., & Peters, M. (2021). Exploring the role of service quality, atmosphere and food for revisits in restaurants by using a e-mystery guest approach. *Journal of Hospitality and Tourism Insights, 4*(3), 351-369. https://doi.org/10.1108/JHTI-04-2020-0048

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: Analyzing text with the natural language toolkit. *O'Reilly Media*.

Blal, I., & Sturman, M. C. (2014). The differential effects of the quality and quantity of online reviews on hotel room sales. *Cornell Hospitality Quarterly, 55*(4), 365-375. https://doi.org/10.1177/1938965514533419

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993-1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Buehring, J., & O'Mahony, B. (2019). Designing memorable guest experiences: Development of constructs and value generating factors in luxury hotels. *Journal of Hospitality and Tourism Insights, 2*(4), 358-376. https://doi.org/10.1108/JHTI-11-2018-0077

Chaudhuri, A., & Holbrook, M. B. (2001). The chain of effects from brand trust and brand affect to brand performance: The role of brand loyalty. *Journal of Marketing, 65*(2), 81-93. https://doi.org/10.1509/jmkg.65.2.81.18255

Chen, Y.-L., Chen, J., Liu, W.-Y., & Sharma, T. (2020). Expected benefits of people interactions and guest experiences. *International Hospitality Review, 34*(2), 187-202. https://doi.org/10.1108/IHR-04-2020-0010

Dang, T. D., & Nguyen, M. T. (2023). Systematic review and research agenda for the tourism and hospitality sector: Co-creation of customer value in the digital age. *Future Business Journal, 9*, 94. https://doi.org/10.1186/s43093-023-00274-5

Dang, T. D., & Nguyen, M. T. (2024). Understanding customer perception and brand equity in the hospitality sector: Integrating sentiment analysis and topic modeling. In *Springer Proceedings in Business and Economics* (pp. 413-425). Springer. https://doi.org/10.1007/978-3-031-49105-4_24

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dominici, G., & Guzzo, R. (2010). Customer satisfaction in the hotel industry: A case study from Sicily. *International Journal of Marketing Studies, 2*(2), 3-12. https://doi.org/10.5539/ijms.v2n2p3

European Parliament and Council. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union, L119*, 1-88. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

Gerber, M. (2023). Customer service management with a people-first hospitality touch. *Hospitality Insights*. Retrieved from https://hospitalityinsights.ehl.edu/customer-service-management-people-first-hospitality-touch

Hay, K. (2024, April 15). The Importance of guest feedback in hospitality. *Benbria*. https://benbria.com/guest-feedback-in-hospitality/

Ho, R. C., Withanage, M. S., & Khong, K. W. (2020). Sentiment drivers of hotel customers: A hybrid approach using unstructured data from online reviews. *Asia-Pacific Journal of Business Administration, 12*(3/4), 237-250. https://doi.org/10.1108/APJBA-09-2019-0192

Hoti, M., & Ajdari, J. (2023). Sentiment analysis using the VADER model for assessing company services based on posts on social media. *South East European Journal of Contemporary Business Issues, 4(*3), 123-135. https://doi.org/10.2478/seeur-2023-0043

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168-177. https://doi.org/10.1145/1014052.1014073

Huang, S., Liang, L. J., & Choi, H. C. (2022). How we failed in context: A text-mining approach to understanding hotel service failures. *Sustainability, 14*(5), 2675. https://doi.org/10.3390/su14052675

Hutto, C. J., & Gilbert, E. E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014*, 216–225.

Iqbal, A., Amin, R., Iqbal, J., Alroobaea, R., Binmahfoudh, A., & Hussain, M. (2022). Sentiment analysis of consumer reviews using deep learning. *Sustainability, 14*(17), 10844. https://doi.org/10.3390/su141710844

Kim, Y.-J., & Kim, H.-S. (2022). The impact of hotel customer experience on customer satisfaction through online reviews. *Sustainability, 14*(2), 848. https://doi.org/10.3390/su14020848

Kwok, L., & Xie, K. L. (2016). Factors contributing to the helpfulness of online hotel reviews: Does manager response play a role? *International Journal of Contemporary Hospitality Management, 28*(10), 2156-2177. https://doi.org/10.1108/IJCHM-03-2015-0107

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing, 80*(6), 69-85. https://doi.org/10.1509/jm.15.0420

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies, 5*(1), 1-167.

Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F. (2017). Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management, 59*, 554-563. https://doi.org/10.1016/j.tourman.2016.08.012

Malik, N., & Bilal, M. (2024). Natural language processing for analyzing online customer reviews: A survey, taxonomy, and open research challenges. *PeerJ Computer Science, 10*(1), 2203. https://doi.org/10.7717/peerj-cs.2203

Mauri, A. G., & Minazzi, R. (2013). Web reviews influence on expectations and purchase intentions of hotel potential customers. *International Journal of Hospitality Management, 34*, 99-107. https://doi.org/10.1016/j.ijhm.2013.02.012

Maroco, A. L., & Maroco, J. (2013). Service quality, customer satisfaction and loyalty. *European Journal of Tourism, Hospitality and Recreation, 4*(3), 118-145. https://www.academia.edu/download/70984701/Service_quality_Customer_Satisfaction_an20211002-27270-1ptqf6h.pdf

Mehraliyev, F., Chan, I. C. C., & Kirilenko, A. P. (2022). Sentiment analysis in hospitality and tourism: A thematic and methodological review. *International Journal of Contemporary Hospitality Management, 33*(8), 2563-2580. https://doi.org/10.1108/IJCHM-02-2021-0132

Nyagadza, B., Mazuruse, G., Muposhi, A., & Chigora, F. (2022). Effect of hotel overall service quality on customers' attitudinal and behavioural loyalty: Perspectives from Zimbabwe. *Tourism Critiques, 3*(1), 42-71. https://doi.org/10.1108/TRC-12-2021-0026

Oberson, M. (2023). How good acoustic design can enhance guest experience and comfort. *Hospitality Insights*. Retrieved from https://hospitalityinsights.ehl.edu/acoustic-design-guest-experience

Ounacer, S., Daif, A., El Ghazouani, M., & Azzouazi, M. (2024). Enhancing hotel services through sentiment analysis. *Engineering Applications of Artificial Intelligence, 6*(3), 429-443. https://doi.org/10.1007/978-3-031-50300-9_24

Ounacer, S., Mhamdi, D., Ardchir, S., Daif, A., & Azzouazi, M. (2023). Customer sentiment analysis in hotel reviews through Natural Language Processing techniques." *International Journal of Advanced Computer Science and Applications, 14*(1), 45-58. https://doi.org/10.14569/IJACSA.2023.0140162

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135. https://doi.org/10.1561/1500000011

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64*(1), 12-40. https://www.marketeurexpert.fr/wp-content/uploads/2023/12/servqual.pdf

Prabhakar, R. (2023). Ethical data practices in hospitality: Protecting guest privacy. *Xerago.* https://www.xerago.com/Industries/Hospitality/guest-data-and-privacy

Puh, K., & Bagić Babac, M. (2022). Predicting sentiment and rating of tourist reviews using machine learning. *Journal of Hospitality and Tourism Insights, 6*(3), 1188-1204. https://doi.org/10.1108/JHTI-02-2022-0078

Ray, A., & Ma, L. (2021). Operational changes and performance outcomes: Analysis on hotels of five Asia–Pacific countries. *International Journal of Global Business and Competitiveness, 16*, 77-88. https://doi.org/10.1007/s42943-021-00041-9

Samodra, D. B., & Sutoyo, E. (2023). Aspect-based sentiment analysis of hotels in Bali on Tripadvisor using BERT algorithm. *International Journal of Advanced Information Systems and Applications, 4*(1), 29-40. https://doi.org/10.25008/ijadis.v4i2.1284

Samy Talaat, A. (2023). Sentiment analysis classification system using hybrid BERT models. *Journal of Big Data, 10*, 110. https://doi.org/10.1186/s40537-023-00781-w

Seyfi, S., Vo-Thanh, T., & Zaman, M. (2024). Hospitality in the age of Gen Z: A critical reflection on evolving customer and workforce expectations. *International Journal of Contemporary Hospitality Management, 36*(13), 118-134. https://doi.org/10.1108/IJCHM-01-2024-0035

Shao, Y., Cho, S., Tse, E.C., & Zou, S. (2017). The effect of front desk employees' characteristics on the first impression of customers and customers' satisfaction. *Food Science (MU)*, 10355. https://doi.org/10.32469/10355/62085

Sutherland, I., Sim, Y., Lee, S. K., Byun, J., & Kiatkawsin, K. (2020). Topic modeling of online accommodation reviews via Latent Dirichlet Allocation. *Sustainability, 12*(5), 1821. https://doi.org/10.3390/su12051821

Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. *Applied Sciences, 13*(7), 4550. https://doi.org/10.3390/app13074550

Tang, L., Wang, X., & Kim, E. (2022). Predicting conversion rates in online hotel bookings with customer reviews. *Journal of Theoretical and Applied Electronic Commerce Research, 17*(4), 1264-1278. https://doi.org/10.3390/jtaer17040064

Troisi, O., Visvizi, A., & Grimaldi, M. (2023). Digitalizing business models in hospitality ecosystems: Toward data-driven innovation. *European Journal of Innovation Management, 26*(7), 1234-1248. https://doi.org/10.1108/EJIM-09-2022-0540

Tuerlan, T., Li, S., & Scott, N. (2021). Customer emotion research in hospitality and tourism: Conceptualization, measurements, antecedents and consequences. *International Journal of Contemporary Hospitality Management, 33*(8), 2741-2772. https://doi.org/10.1108/IJCHM-11-2020-1257

Ved, M. (2020, August 20). Text preprocessing for NLP and machine learning tasks - *Experfy Insights*. https://resources.experfy.com/ai-ml/text-preprocessing-for-nlp-and-machine-learning-tasks/

Vinodhini, G., & Chandrasekaran, R. M. (2012*).* Sentiment analysis and opinion mining: A survey. *International Journal of Advanced Research in Computer Science and Software Engineering, 2*(6), 282-292. Available online at: https://www.researchgate.net/profile/Vinodhini-G-2/publication/265163299_Sentiment_Analysis_and_Opinion_Mining_A_Survey/links/54018f330cf2bba34c1af133/Sentiment-Analysis-and-Opinion-Mining-A-Survey.pdf

Wolf, T., et al. (2020). *Transformers: State-of-the-art natural language processing*. https://github.com/huggingface/transformers

Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management, 44,* 120-130. https://doi.org/10.1016/j.ijhm.2014.10.013

Yang, Y., Mao, Z., & Tang, J. (2017). Understanding guest satisfaction with urban hotel location. *Journal of Travel Research, 56*(6), 710-725. https://doi.org/10.1177/0047287517691153

Yrjölä, M., Rintamäki, T., Saarijärvi, H., Joensuu, J., & Kulkarni, G. (2019). A customer value perspective to service experiences in restaurants. *Journal of Retailing and Consumer Services, 51*, 91-101. https://doi.org/10.1016/j.jretconser.2019.05.030

Yu, Y. (2016). Aspect-based sentiment analysis on hotel reviews. *Stanford University*. https://cs229.stanford.edu/proj2016spr/report/032.pdf

Zarezadeh, Z. Z., Rastegar, R., & Xiang, Z. (2022). Big data analytics and hotel guest experience: A critical analysis of the literature. *International Journal of Contemporary Hospitality Management, 34*(6), 2320-2336. https://doi.org/10.1108/IJCHM-10-2021-1293

Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022). A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *arXiv:2203.01054*. https://doi.org/10.48550/arXiv.2203.01054

Zhang, Z., Ye, Q., Law, R., & 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), 694-700. https://doi.org/10.1016/j.ijhm.2010.02.002

Zvornicanin, E. (2024). When coherence score is good or bad in topic modeling? *Baeldung on Computer Science*. Retrieved from https://www.baeldung.com/cs/topic-modeling-coherence-score

Zygiaris, S., Hameed, Z., Alsubaie, M. A., & Rehman, S. U. (2022). Service quality and customer satisfaction in the post-pandemic world: A study of Saudi auto care industry. *Frontiers in Psychology, 13*, 842141. https://doi.org/10.3389/fpsyg.2022.842141