BEMM466 Business Project

Title: Comparative Sentiment Analysis of Singapore Airlines' Customer Reviews During Pre-COVID, COVID-19, and Post-COVID Periods Based on TripAdvisor Reviews

Natamon Tosawat

Student ID : 730102604

Supervisor: Dr Stuart So

MSc Business Analytics, Year 2024/2025

OneDrive link	BEMM466_Sentiment_Analysis_Natamon
GitHub link	https://github.com/natamontos/BEMM466_sentiment_analysis_Natamon
Data source	https://www.tripadvisor.com/ShowUserReviews-g1-d8729151-r942167897-Singapore_Airlines-World.html
	https://www.kaggle.com/datasets/kanchana1990/singapore-airlines-reviews

Executive Summary

Singapore Airlines (SIA) has consistently ranked among the top three airlines in the Skytrax World Airline Awards for over two decades. These awards, considered the "Oscars of the aviation industry," evaluate airlines using a global survey methodology covering service quality, comfort, and overall satisfaction (International Airport Review, 2024). High rankings enhance brand reputation and influence customer trust (Agustia et al., 2020). User-generated reviews on TripAdvisor complement these awards by providing real-time insights into customer perceptions (Filieri, 2015). This study analyzes 10,000 SIA reviews spanning pre-COVID, during COVID-19, and post-COVID periods, leveraging sentiment and thematic analysis to uncover evolving customer priorities.

This study aims to explore customer sentiments and themes in Singapore Airlines reviews, identifying satisfaction drivers over three critical periods: pre-COVID, during COVID-19, and post-COVID. Using DistilBERT for sentiment analysis, BERTopic for thematic mapping, and LSTM for forecasting, the research benchmarks insights against Skytrax criteria. These objectives support SIA in refining services, anticipating future trends, and maintaining competitive leadership in a post-pandemic market. The findings also offer a framework for broader aviation industry strategies during external disruptions.

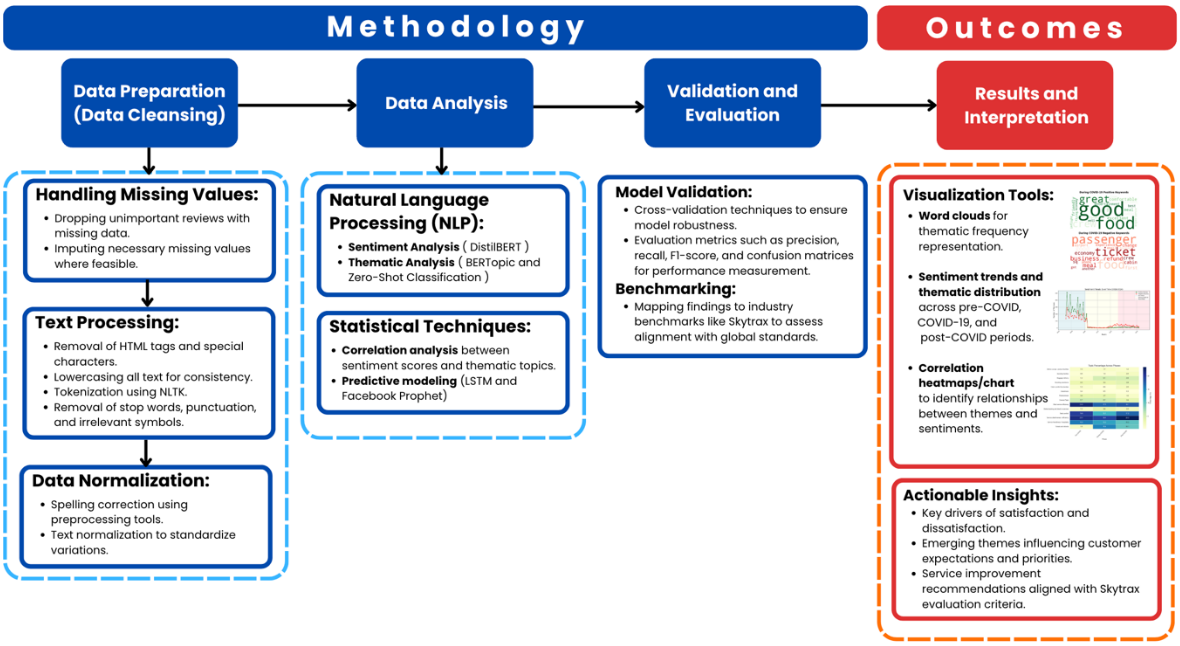


Figure 1: Graphical Abstract for Research

According to Figure 1, the methodology integrates advanced NLP and statistical techniques to analyze 10,000 TripAdvisor reviews of Singapore Airlines. Text preprocessing ensures data consistency through tokenization, lemmatization, and removal of irrelevant elements. Sentiment analysis leverages DistilBERT for precise classification, while BERTopic and Zero-Shot Classification extract thematic insights. Statistical modeling, including LSTM and Prophet, forecasts customer feedback trends with high accuracy, evaluated using RMSE and MAPE. Findings are benchmarked against Skytrax criteria, providing actionable insights to enhance customer satisfaction and operational efficiency.

The analysis of SIA customer reviews reveals significant insights into customer satisfaction trends across three distinct phases: pre-COVID, during COVID-19, and post-COVID. Sentiment analysis using DistilBERT identified 56.25% of reviews as positive and 43.75% as negative. Thematic analysis with BERTopic and Zero-Shot Classification refined these insights, uncovering key drivers of satisfaction and dissatisfaction.

Positive sentiment was predominantly linked to Service attentiveness/efficiency and Meal service efficiency. These attributes consistently received favorable reviews across all phases, affirming SIA's operational excellence and attentive in-flight service. Seat comfort emerged as another strong contributor, particularly in premium travel classes. During the pandemic, customer appreciation for enhanced health protocols and cleanliness also stood out, emphasizing the importance of safety measures in maintaining trust.

Conversely, negative feedback highlighted persistent challenges in Tickets and refunds and Seat comfort. The pandemic amplified dissatisfaction with ticketing processes, as reflected in the surge of refund-related complaints during 2020-2021. While this negative trend showed improvement in the post-COVID phase, it remains an area for strategic focus. Additionally, discomfort with economy-class seating was frequently cited, particularly as travel volumes normalized.

Temporal analysis revealed shifting customer priorities, with a dramatic increase in reviews focused on operational aspects like ticketing and refunds during COVID-19. Post-COVID, themes such as in-flight service and meal quality regained prominence. Forecasting insights using LSTM models predict a sustained positive trajectory for service efficiency and meal quality but underscore the need for addressing procedural inefficiencies and seating-related concerns to meet evolving customer expectations.

These findings underscore the critical role of personalized service, operational excellence, and proactive issue resolution in driving customer satisfaction.

To address evolving customer needs and sustain its competitive edge, Singapore Airlines should adopt strategies that prioritize financial, psychological, and service-related concerns in the post-COVID landscape.

First, flexible booking policies must be implemented to reduce perceived financial risks. Refundable tickets, fee waivers for changes or cancellations, and travel insurance covering pandemic-related disruptions are critical measures. Immediate and tangible compensations, such as monetary refunds, upgrades, or complimentary services, can significantly alleviate dissatisfaction, particularly for full-service airlines like SIA, where passenger expectations are higher.

Second, digital transformation should remain a priority. Seamless digital platforms for ticketing, refund processes, and real-time communication can build customer trust and operational efficiency. Advanced analytics tools, powered by real-time data, would streamline decision-making and enable proactive responses to disruptions, minimizing uncertainties and enhancing transparency (Pereira, 2023; Song et al., 2020).

Third, addressing seat comfort concerns is essential. The rise in negative feedback post-COVID highlights the need for cabin redesigns that emphasize personal space and ergonomic seating. Flexible cabin configurations catering to diverse customer needs, including enhanced seating options, would mitigate dissatisfaction and improve the in-flight experience.

Finally, investment in staff training is crucial to maintain high standards in Service attentiveness/efficiency and Meal service quality. Personalized interactions and consistent meal delivery reinforce SIA’s premium reputation and drive positive customer sentiment.

By adopting these targeted measures, SIA can navigate post-pandemic challenges, foster customer loyalty, and uphold its position as an industry leader.

This study offers critical insights into the evolving dynamics of customer satisfaction, particularly in the context of post-pandemic recovery for the aviation industry. By leveraging advanced analytical techniques such as sentiment analysis, thematic modeling, and predictive forecasting, the research provides actionable intelligence to address key challenges and opportunities for Singapore Airlines (SIA).

Aligning customer feedback with Skytrax evaluation criteria ensures that the findings are not only industry-relevant but also benchmarked against global standards of excellence. The study's insights into service strengths, such as Service attentiveness/efficiency and Meal service quality, provide a roadmap for sustaining SIA's premium reputation. Simultaneously, addressing challenges such as ticketing, refunds, and seat comfort enables the airline to enhance operational resilience and customer satisfaction.

The integration of predictive tools equips SIA to anticipate future trends, enabling proactive adaptations to shifting customer needs. Ultimately, this research sets a benchmark for data-driven strategies that ensure long-term competitiveness and industry leadership.

In conclusion, this research highlights the critical role of customer feedback in shaping service strategies for Singapore Airlines amidst a rapidly evolving aviation landscape. By analyzing customer sentiments across key phases—pre-COVID, during COVID, and post-COVID—the study identifies actionable insights for addressing customer expectations, enhancing satisfaction, and mitigating dissatisfaction. The findings emphasize the need for flexible policies, digital transformation, and targeted service enhancements to maintain SIA's premium positioning. Furthermore, predictive insights enable the airline to anticipate emerging trends and proactively adapt to market demands. These outcomes provide a strategic framework for SIA to uphold its competitive edge and deliver exceptional customer experiences.

Table of Contents

Introduction

Background

Singapore Airlines (SIA) has consistently ranked among the top three airlines in the Skytrax World Airline Awards for over two decades. These awards, often referred to as the "Oscars of the aviation industry," are widely recognized as a benchmark of excellence due to their rigorous global survey methodology and comprehensive evaluation criteria (International Airport Review, 2024). The Skytrax assessment spans categories such as service quality, comfort, in-flight experience, and overall customer satisfaction. High rankings not only enhance an airline's reputation but also foster consumer trust, influencing passenger decision-making significantly (Agustia et al., 2020). Airlines frequently use these benchmarks to refine their services and align closely with evolving customer expectations (Samah et al., 2022).

In addition to formal awards, user-generated reviews on platforms like TripAdvisor provide critical, real-time feedback and first-hand customer experiences (Tripadvisor, 2024). These reviews capture unfiltered, spontaneous sentiments, making them invaluable for understanding customer satisfaction and identifying service improvement opportunities (Filieri, 2015). Unlike structured surveys, these reviews offer authenticity, shaping consumer perceptions and influencing travel decisions (Abubakar & Ilkan, 2016; Bigne et al., 2019; Lou & Yuan, 2018). Leveraging such insights enables airlines to adapt effectively to changing customer needs.

The COVID-19 pandemic has dramatically reshaped the airline industry, altering customer priorities and operational dynamics. This study analyzes 10,000 anonymized TripAdvisor reviews of SIA collected between August 5, 2018, and March 12, 2024, across three key periods: pre-COVID, during COVID-19, and post-COVID. By applying sentiment analysis using DistilBERT, thematic analysis with BERTopic, and forecasting using LSTM models, the research identifies key trends and shifts in customer perceptions during these critical phases. Comparing sentiment-driven insights with Skytrax evaluation criteria, the study provides actionable recommendations to help SIA maintain its competitive edge and adapt to evolving market demands in a post-pandemic world.

Problem Statement

Singapore Airlines, a global leader in the aviation industry, operates in an environment where customer satisfaction is a critical determinant of success. The COVID-19 pandemic has profoundly disrupted the industry, significantly altering customer behaviors and expectations (Das et al., 2022). These changes have heightened the need for data-driven insights to adapt service offerings and maintain a competitive edge. Despite its longstanding reputation for excellence, the unstructured nature of customer reviews on platforms like TripAdvisor presents challenges in systematically deriving actionable insights to enhance service quality and align with customer priorities (Deasie, Inc., 2023).

This study seeks to address the difficulty of analyzing customer feedback to uncover meaningful insights. By employing sentiment analysis and thematic analysis, the research bridges the gap between raw data and actionable intelligence. Focusing on feedback from pre-COVID, during COVID-19, and post-COVID periods, the study aims to identify evolving customer expectations and key drivers of satisfaction and dissatisfaction. These findings will guide Singapore Airlines in refining service strategies to remain at the forefront of the industry.

Project Objective

This research aims to explore and understand the evolving sentiments and themes reflected in customer reviews of Singapore Airlines. By utilizing advanced analytical techniques such as sentiment analysis and thematic analysis, the study seeks to provide actionable insights into how customer satisfaction drivers have changed over time, particularly in response to significant disruptions such as the COVID-19 pandemic. Additionally, the research will examine the alignment between customer feedback and global benchmarks like the Skytrax evaluation criteria to ensure that the findings are grounded in industry standards and contribute meaningfully to service enhancements.

Through this comprehensive approach, the study aims to support Singapore Airlines in refining its service offerings, anticipating future trends in customer preferences, and sustaining its position as an industry leader in the rapidly changing aviation market. The insights derived would not only inform strategic decision-making but also set a benchmark for integrating customer feedback into the broader airline industry's efforts to navigate external disruptions effectively. Moreover, the study employs time-series analysis to forecast future trends in customer feedback, enabling Singapore Airlines to anticipate emerging service needs and proactively adapt its strategies. By integrating these insights into decision-making processes, the research aims to develop evidence-based recommendations that enhance customer experience, foster loyalty, and maintain the airline's competitive advantage in a rapidly evolving market.

Significance and Motivation of the Research

The significance of this research lies in its ability to provide Singapore Airlines with actionable insights into customer sentiments during a time of unprecedented change. By leveraging advanced methodologies, the study addresses the challenges posed by unstructured feedback and highlights strategies for maintaining excellence in customer service.

Firstly, the research would enable Singapore Airlines to adapt its offerings to meet evolving customer needs, particularly in areas such as health, safety, and flexibility. These findings will be instrumental in fostering loyalty and ensuring customer satisfaction in a post-pandemic landscape.

Secondly, aligning insights with Skytrax evaluation criteria ensures that the airline benchmarks its performance against industry standards while identifying opportunities for differentiation. This alignment strengthens Singapore Airlines' ability to maintain its competitive edge while adhering to global benchmarks of service quality.

Beyond Singapore Airlines, this research offers broader industry implications. It provides a framework for understanding how external disruptions, such as the COVID-19 pandemic, influence customer satisfaction, enabling airlines to anticipate and respond to emerging challenges effectively. The forecasting of sentiment trends further equips stakeholders with the foresight needed to navigate a rapidly changing market.

Finally, the motivation for this study stems from the growing importance of data-driven decision-making in the aviation industry. By employing cutting-edge tools like DistilBERT and BERTopic, this research sets a benchmark for utilizing advanced methodologies to enhance customer understanding and foster innovation in service delivery.

Research Questions

1. How has customer sentiment evolved across the periods of pre-COVID, during COVID-19, and post-COVID, based on the analysis of customer reviews, and what are the key insights from the binary sentiment classification into positive and negative categories?
2. What themes and topics are strongly associated with positive and negative sentiments, and how do these themes vary across different time periods?
3. How can the themes identified through thematic analysis and aligned with Skytrax evaluation criteria be prioritized to enhance Singapore Airlines' service offerings and sustain its industry-leading position?
4. How can trends in the prevalence of main topics within positive and negative sentiments be used to forecast changes in customer feedback and anticipate future service priorities?

Literature Review

Overview of Sentiment Analysis in the Airline Industry

Understanding customer feedback through sentiment analysis is critical for identifying factors that drive satisfaction and dissatisfaction in service industries like airlines (Komischke, 2024). Traditional sentiment analysis tools, such as VADER and TextBlob, are widely recognized for their simplicity, ease of implementation, and ability to provide quick insights into customer sentiment (Rafalski, 2024). VADER, a rule-based model, excels in analyzing sentiment in short and informal texts, such as social media posts and user reviews, by effectively handling features like emoticons, acronyms, and exclamation points (Aryan, 2024; Frenzel, 2024; Youvan, 2024). Similarly, TextBlob, a lexicon-based model, offers intuitive sentiment polarity scoring and is particularly effective for texts with clear and explicit expressions of sentiment (Afaf, 2024). Although both tools face challenges in capturing subtle nuances like sarcasm or complex sentence structures (Kaur et al., 2024), their lightweight and computationally efficient nature makes them highly attractive for exploratory sentiment analysis and as benchmarks against advanced models like DistilBERT.

Advanced transformer-based models, such as BERT (Bidirectional Encoder Representations from Transformers), have been developed to address these limitations by providing deep contextual understanding (Pontes, 2024). However, BERT's computational intensity makes it less suitable for large datasets. DistilBERT, a lightweight version of BERT, retains 97% of BERT's performance while being 60% faster and using 40% fewer resources (Sanh et al., 2019). This efficiency makes DistilBERT particularly suitable for analyzing extensive datasets like the 10,000 TripAdvisor reviews in this study, allowing for accurate sentiment classification into positive and negative categories. The binary classification approach highlights actionable insights by focusing on areas where service improvements can be prioritized.

Gaps and Controversies in Sentiment Analysis

While sentiment analysis has advanced significantly, notable challenges remain in its effective application within the airline industry. Transformer-based models like DistilBERT excel in understanding context-rich text but face limitations such as high computational costs and difficulty in handling multilingual datasets and cultural nuances (Naveen & Trojovský, 2024). For example, idiomatic expressions like "the flight was a rollercoaster" or sarcasm such as "just loved the 2-hour delay" can be misinterpreted without contextual knowledge (Ramalingam et al., 2023; Dorssers, 2024). Lexicon-based methods like VADER, while efficient, often fail to identify such complexities, particularly in mixed-sentiment reviews (Garcia & Wang, 2020).

Sarcasm detection has shown promise when combined with sentiment information, enhancing accuracy in informal contexts (Majumder et al., 2019). This highlights the potential of hybrid approaches that integrate lexicon-based methods for shorter texts and transformer models for more nuanced reviews. Fine-tuning transformer models with domain-specific datasets can further improve their ability to detect cultural subtleties and emotional depth (Dorssers, 2024).

This study evaluates VADER and DistilBERT to identify the optimal model for analyzing customer reviews. While multilingual considerations are less relevant for the English-only dataset, models like XLM-RoBERTa offer a solution for future cross-lingual applications. XLM-RoBERTa, trained on 100 languages, has demonstrated superior performance in tasks involving low-resource languages, making it a valuable tool for extending sentiment analysis across diverse customer bases (Huggingface, 2024). Its ability to handle nuanced multilingual contexts without sacrificing accuracy positions it as a strong candidate for global airline datasets.

The integration of sentiment analysis into airline operations has become increasingly vital. Negative reviews can amplify dissatisfaction and damage brand reputation, particularly through social media (FareTrack, 2024). Monitoring customer sentiment in real-time enables airlines to address issues proactively, such as resolving operational problems like delays or poor in-flight service, thus reducing customer churn and protecting revenue. Additionally, analyzing competitor sentiment provides opportunities to attract dissatisfied passengers through targeted marketing strategies, strengthening market position (FareTrack, 2024).

In summary, sentiment analysis in the airline industry requires a balance between efficiency and depth. While transformer models like DistilBERT are effective for nuanced reviews, combining them with simpler methods can optimize performance. Future applications, such as incorporating multilingual frameworks or competitor sentiment insights, can offer actionable strategies to enhance customer satisfaction and operational decision-making.

Forecasting Trends in Customer Sentiments and Themes

Time series forecasting plays a critical role in anticipating trends in customer feedback and planning strategic interventions. Various forecasting methods are available, each with its strengths and limitations. SARIMA (Seasonal Autoregressive Integrated Moving Average) is a widely recognized statistical model for forecasting seasonal time series data. In 2023, Chauhan states that while effective for datasets with clear seasonal and trend patterns, SARIMA requires meticulous parameter tuning and struggles with datasets exhibiting abrupt changes, such as those influenced by the COVID-19 pandemic. Similarly, the ETS (Error, Trend, Seasonality) model provides an interpretable approach by directly modeling error, trend, and seasonality components. Although simpler to implement, ETS is less effective for datasets with complex, non-linear patterns (Fatima & Rahimi, 2024).

In contrast, LSTM (Long Short-Term Memory) networks and Facebook Prophet provide more flexible and adaptive forecasting solutions. LSTM networks, a type of recurrent neural network, excel in capturing non-linear dependencies and long-term trends, making them suitable for dynamic datasets (Rao, 2024). According to Damle (2024), Facebook Prophet is a user-friendly tool designed for forecasting time series data with irregularities, seasonal effects, and missing values. Its ease of implementation and interpretability make it particularly advantageous for analyzing customer sentiment and thematic trends.

This study employs LSTM and Facebook Prophet to forecast the prevalence of main topics associated with customer sentiments. These methods were selected for their ability to adapt to non-linear and irregular data patterns, providing insights into emerging trends in customer feedback and service priorities. Their complementary strengths make them a more suitable choice than SARIMA or ETS for analyzing data influenced by structural breaks and disruptions like the pandemic.

Thematic Analysis with BERTopic and Zero-Shot Classification

Thematic analysis is an essential tool for uncovering deeper insights into customer experiences. Traditional methods, such as Latent Dirichlet Allocation (LDA), are widely used for topic modeling but often fail to capture contextual nuances, especially in reviews containing diverse themes (Kaur et al., 2024). BERTopic addresses these limitations by leveraging BERT embeddings to identify contextually relevant themes with high accuracy. Grootendorst in 2022 remarks that its ability to dynamically adjust the number of topics without requiring pre-defined parameters makes it ideal for analyzing diverse and unstructured user-generated content.

Despite its strengths, BERTopic faces limitations when reviews contain multiple overlapping themes, such as food, seating, and service quality, within a single review. To overcome this, Zero-Shot Classification is employed as a complementary approach. Zero-Shot Classification uses large transformer models to assign the most relevant topic to reviews with overlapping themes (Moayeri et al., 2024), ensuring a more granular understanding of customer feedback. This dual approach enables the mapping of identified themes to the Skytrax evaluation criteria, allowing for benchmarking and deeper insights into why Singapore Airlines consistently ranks among the top three airlines.

Thematic trends are further analyzed across three critical phases—pre-COVID, during COVID-19, and post-COVID—to explore how customer experiences and priorities have shifted. This comprehensive thematic analysis provides actionable insights into areas for service improvement and strategic planning.

Debates and Future Directions

While sentiment analysis and thematic modeling offer powerful tools for understanding customer experiences, debates remain regarding their reliability and interpretability. For instance, tools like VADER and TextBlob are often criticized for their inability to handle nuanced language or contextual sarcasm, which are common in customer reviews (Kaur et al., 2024). Similarly, while BERTopic provides high accuracy in thematic modeling, its reliance on pretrained embeddings may introduce biases if the underlying data is not representative of the target audience (Grootendorst, 2022).

Future research should focus on enhancing the multilingual capabilities of sentiment analysis models to address the diverse linguistic backgrounds of airline customers. Additionally, integrating cultural context into thematic analysis could improve the interpretability of insights for global audiences. Finally, combining sentiment analysis with behavioral data, such as booking trends or service complaints, could provide a more holistic view of customer satisfaction drivers, enabling airlines to make more informed strategic decisions.

Research Design

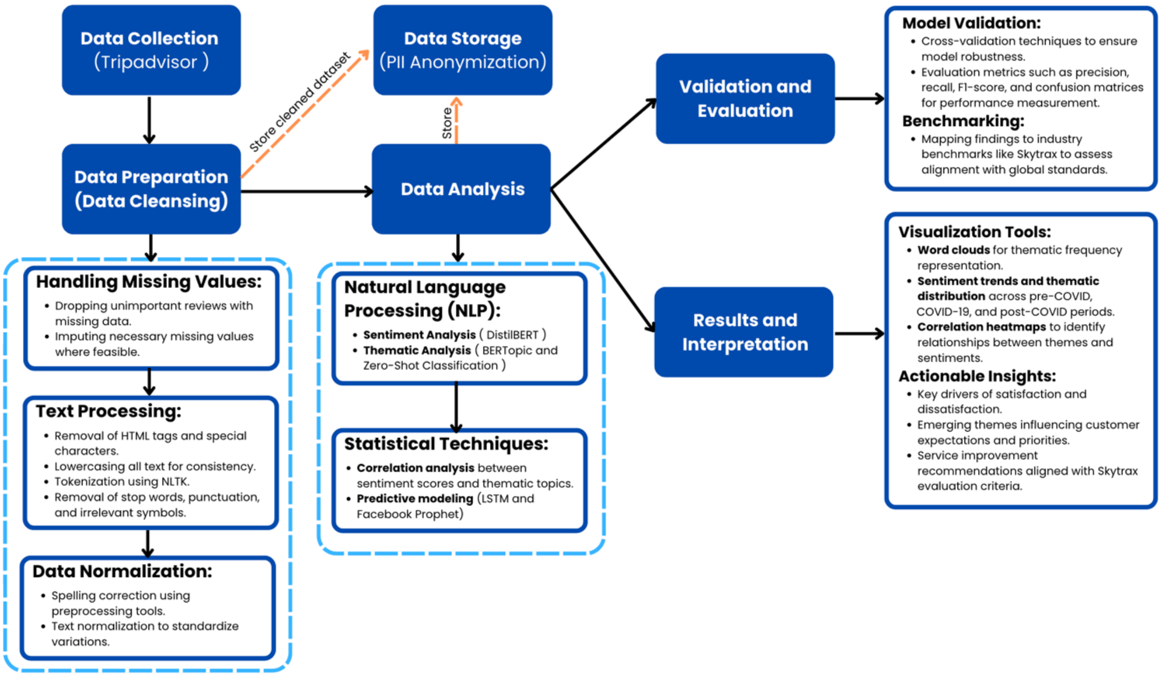


Figure 2: Research Framework and Approach

The research design for this study systematically analyzes and forecasts customer sentiments expressed in TripAdvisor reviews for Singapore Airlines, as shown in the framework. The dataset, comprising 10,000 anonymized reviews spanning 2018-2024, was collected from TripAdvisor and published in CSV format on Kaggle, ensuring compliance with ethical standards through PII anonymization.

Data preparation involves tokenization, normalization, and handling missing values to create a clean dataset. Sentiment analysis using DistilBERT classifies reviews into positive and negative categories, while thematic analysis with BERTopic and Zero-Shot Classification identifies key themes. Correlation analysis examines relationships between sentiments and themes, while time-series forecasting with LSTM and Facebook Prophet predicts future trends.

The results, visualized via word clouds, sentiment trends, and heatmaps, inform actionable insights for service improvement aligned with Skytrax evaluation criteria.

Methodology

Key Theories and Concepts

The selection of methodologies in this research is guided by the need to analyze complex and unstructured customer reviews while ensuring actionable insights. This study employs advanced Natural Language Processing (NLP) tools, machine learning models, and statistical techniques to address the objectives comprehensively.

Data Preparation

Preprocessing ensures data consistency and quality, using the Natural Language Toolkit (NLTK) for tokenization, stop-word removal, and normalization (Bird et al., 2009). These steps minimize noise and optimize data for analysis.

Sentiment Analysis

Sentiment analysis plays a fundamental role in this study by identifying customer emotions and sentiment polarity. DistilBERT, a transformer-based model, predicts sentiment by capturing contextual relationships between words, offering high accuracy and computational efficiency (Sanh et al., 2019). In contrast, VADER (Valence Aware Dictionary and Sentiment Reasoner) is optimized for informal text, incorporating slang, abbreviations, and emoticons in its lexicon while applying rules for linguistic nuances like negation and intensifiers (Hutto & Gilbert, 2014). Though effective for capturing general sentiment in short texts, VADER may struggle with the complex syntactic structures common in detailed customer reviews, where deep learning models like DistilBERT excel (Devlin et al., 2019).

Table 1: DistilBERT Model Equations (Sanh et al., 2019):

DistilBERT Model Equations	Description
$$P(y\|x) = \frac{e^{f(x)}}{\sum_i e^{f_i(x)}}$$ Where: • $P(y\|x)$: Probability of class y given input x. • $f(x)$: Logit or score for class y computed from the model. • $\sum_i e^{f_i(x)}$: Sum of exponentials of all logits for all classes i, used for normalization. • $e^{f(x)}$: Exponential of the logit for class y.	DistilBERT applies a softmax function over sentiment scores to predict polarity efficiently while maintaining high accuracy.

$$\text{Compound Score} = \frac{\sum(\text{valence of each token} \times \text{adjustment factor})}{\sqrt{\sum(\text{valence of each token}^2)}}$$ Where: • **Valence of each token**: Sentiment score of a word or token. • **Adjustment factor**: Modifier for context (e.g., negation, emphasis). • $\sum$: Summation of all weighted sentiment values. • $\sqrt{\sum(\text{valence of each token}^2)}$: Normalization factor. • **Compound Score**: Overall sentiment score (-1 to +1).	VADER calculates a normalized compound score by summing weighted lexicon-based sentiment values, effectively capturing sentiment trends in informal and expressive text.

To evaluate the model's performance, metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC were employed, ensuring a comprehensive assessment across classification thresholds.

Thematic Analysis and Topic Modeling

Thematic analysis aims to uncover recurring themes and topics in customer reviews. BERTopic leverages BERT embeddings to dynamically extract contextually relevant themes, addressing the limitations of Latent Dirichlet Allocation (LDA), which often fails to manage overlapping topics (Grootendorst, 2022; Blei et al., 2003). To enhance granularity, Zero-Shot Classification is employed to assign the most relevant topic to each review, even for those with overlapping or ambiguous themes.

Table 2: BERTopic Model Equations (Grootendorst, 2022):

BERTopic Model Equations	Description
$$P(z\|w) = \frac{P(w\|z)P(z)}{P(w)}$$ Where: • $P(z\|w)$: Probability of topic z given word w. • $P(w\|z)$: Probability of word w in topic z. • $P(z)$: Prior probability of topic z. • $P(w)$: Probability of word w.	BERTopic uses BERT embeddings and Bayes' rule to dynamically extract and model topics from unstructured text.

Table 3: Zero-Shot Classification Model Equations (Yin et al., 2019):

Zero-Shot Classification Model Equations	Description
$$P(t\|x) = \text{argmax}_{t \in T}\ P(x\|t)$$ Where: • $P(t\|x)$: Probability of topic t given input x. • $\text{argmax}_{t \in T}$: The topic t within the set of all topics T that maximizes the probability. • $P(x\|t)$: Probability of input x being associated with topic t. • T: Set of all possible topics.	Zero-Shot Classification leverages pre-trained transformer models to assign topics without requiring domain-specific training.

Forecasting Trends

Time-series forecasting enables this study to predict future trends in customer feedback with high accuracy. To ensure robust predictions, evaluation metrics such as RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error) were employed to assess model performance and identify the most suitable forecasting model. Based on these evaluations, the model with superior accuracy was selected for use in this study.

Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, excel in capturing non-linear dependencies and temporal patterns, making them well-suited for dynamic datasets like customer feedback (Hochreiter & Schmidhuber, 1997). Meanwhile, Facebook Prophet offers robustness in handling irregularities, seasonal trends, and missing data, making it effective for broader temporal patterns (Taylor & Letham, 2018). However, only the model demonstrating higher accuracy in evaluations was used for forecasting in this research.

Table 4: LSTM Model Equations (Hochreiter & Schmidhuber, 1997):

LSTM Model Equations	Description
$$h_t = \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$ $$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$ $$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o)$$ $$h_t = o_t \odot \tanh(c_t)$$ Where: • h_t : Hidden state at time t. • c_t : Cell state at time t. • x_t : Input at time t.	LSTM networks capture temporal dependencies in sequential data by maintaining long-term and short-term memory through the interaction of hidden states (h_t) and cell states (c_t).

f_t , i_t , o_t: Forget, input, and output gates, respectively. W_{xh}, W_{hh} W_{xo} W_{ho} Weight matrices for input and hidden states. b_h , b_o : Bias terms. σ : Sigmoid activation function. $\odot$: Element-wise multiplication.	

Table 5: Facebook Prophet Model Equations (Taylor & Letham, 2018):

Facebook Prophet Model Equations	Description
$$y_t = g(t) + s(t) + h(t) + \epsilon_t$$ Where: y_t : Observed value at time t. $g(t)$: Trend component modeling long-term changes in data. $s(t)$: Seasonal component capturing periodic patterns. $h(t)$: Holiday effects accounting for specific events. ϵ_t : Error term representing random noise.	Facebook Prophet models time-series data by combining trend, seasonality, and holiday effects, providing flexibility to handle irregularities, missing values, and outliers.

COVID-19 was incorporated as a pivotal external variable in this study, serving as a temporal marker to segment the dataset into Pre-COVID, During-COVID, and Post-COVID phases. The significant decline in global flight frequencies, as shown in Figure 3, provided a critical context for understanding shifts in customer sentiment. The pandemic resulted in widespread cancellations, refund requests, and heightened dissatisfaction with refund policies, such as those of Singapore Airlines, which impose restrictions on non-refundable tickets and travel-agent bookings (Singapore Airlines, 2024). These disruptions served as key inputs for analyzing how operational challenges influence customer experiences.

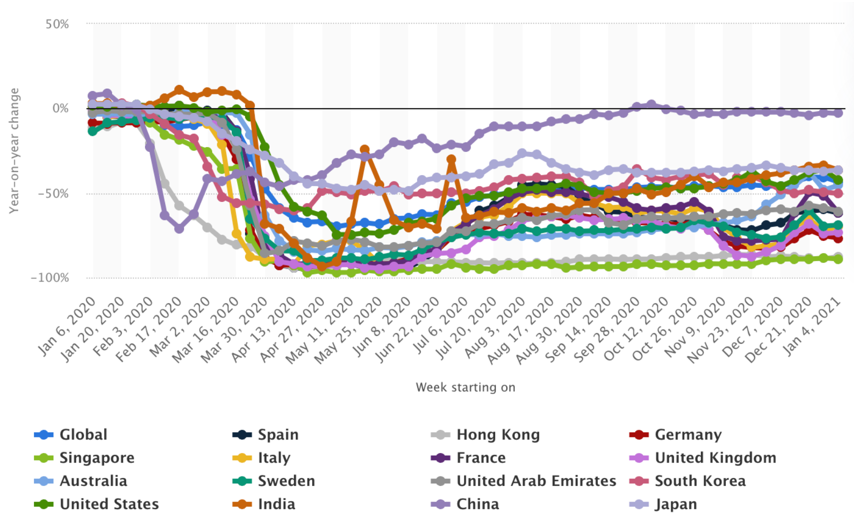


Figure 3: Year-on-year change of weekly flight frequency of global airlines (Statista, 2023)

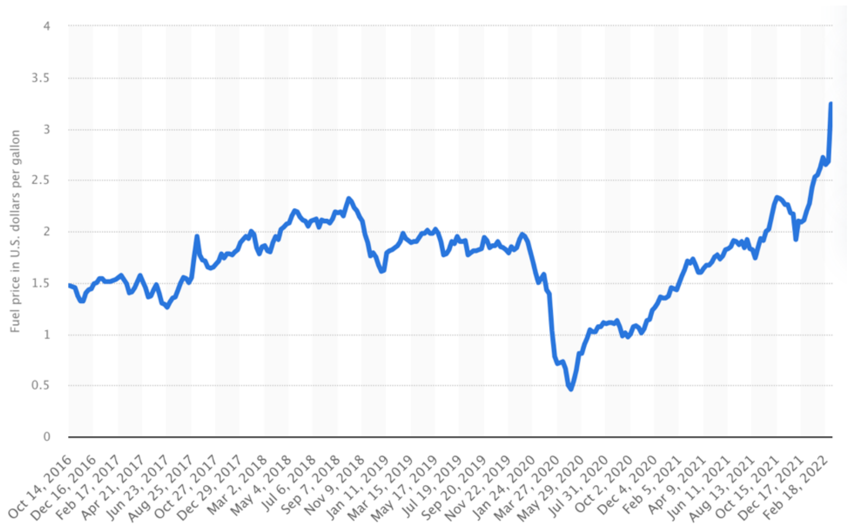


Figure 4: Average weekly jet fuel price (Statista, 2024)

Economic indicators, such as jet fuel prices and inflation rates, were also integrated into the analysis to identify their influence on customer satisfaction. For example, the sharp rise in jet fuel prices post-COVID, as illustrated in Figure 4, likely contributed to increased ticket prices and corresponding negative sentiment. Similarly, inflation and reduced disposable income were considered to evaluate their impact on customer travel behavior and satisfaction trends (Statista, 2024). These economic variables were essential for capturing broader market dynamics and their role in shaping customer sentiment.

Competitive dynamics were another critical consideration in the forecasting methodology. Airlines introduced innovations such as ultra-long-haul flights, designed to minimize layovers and reduce passenger exposure to crowded airports, which could have enhanced customer satisfaction during the pandemic (Bauer et al., 2020). By incorporating these external factors into the forecasting process, the study aimed to generate actionable insights for identifying trends and strategic opportunities.

The integration of these external variables allowed for a comprehensive analysis of customer sentiment trends, contextualizing them within real-world disruptions. These insights were further used to identify patterns that could inform actionable strategies. This methodological approach ensures that sentiment forecasting captures not only historical trends but also dynamic external influences, enhancing its practical applicability.

Data Validation and Cross-Validation Strategies

To ensure the reliability and robustness of the results, this study incorporates data validation and cross-validation techniques:

Handling Missing Values

$$\text{Missing Percentage} = \frac{\text{Number of Missing Values}}{\text{Total Number of Values}} \times 100$$

Missing values can lead to biased results or reduced model performance. Columns with a missing value percentage exceeding a predefined threshold (e.g., 30%) were excluded. For others, imputation was performed using methods such as mean or mode replacement, or by dropping rows if appropriate.

Duplicate Detection and Removal

$$D_{\text{duplicates}} = \sum_{i=1}^{N} I(x_i = x_j) \quad \text{where } i \neq j$$

Duplicates in the dataset inflate counts or biases the analysis. D duplicates represents the count of duplicate rows in the dataset. Duplicates were removed to ensure data uniqueness.

Normalization of Numerical Data

$$X_{\text{scaled}} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

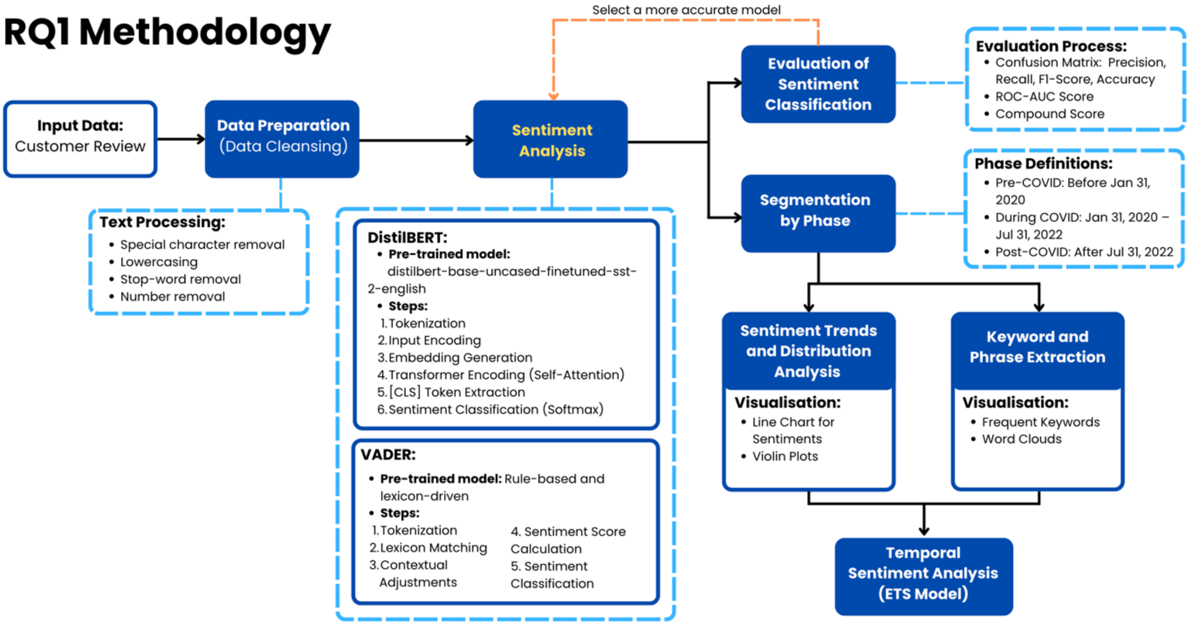
To standardize numerical fields, Min-Max scaling was employed, ensuring the data values lie within a defined range (e.g., [0, 1]). This process prevents numerical columns with larger ranges from dominating the analysis.

Train-Test Split

$$\text{Train-Test Split} = \frac{\text{Training Data Size}}{\text{Total Data Size}}$$

The dataset is divided into training (80%) and testing (20%) subsets. This method provides an unbiased evaluation of the model's performance on unseen data.

<u>**Data Analysis Procedures**</u>



DistilBERT

```python
from transformers import DistilBertTokenizer, DistilBertForSequenceClassification
import torch
import pandas as pd
import torch.nn.functional as F  # For softmax

# Define device
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")

# Load tokenizer
tokenizer = DistilBertTokenizer.from_pretrained('distilbert-base-uncased')

# Load model
model = DistilBertForSequenceClassification.from_pretrained(
    'distilbert-base-uncased-finetuned-sst-2-english'
)
model.to(device)

# Define classify_sentiment function
def classify_sentiment_and_score(review):
    """
    Classify sentiment and calculate compound score for a given review.

    Parameters:
        review (str): The review text to analyze.

    Returns:
        tuple: Sentiment label (positive/negative) and Compound Score.
    """
    # Tokenize and prepare input
    inputs = tokenizer(review, return_tensors="pt", truncation=True, padding=True).to(device)

    with torch.no_grad():
        outputs = model(**inputs)  # Get logits from the model
        logits = outputs.logits  # Extract raw predictions

        # Apply softmax to calculate probabilities
        probabilities = F.softmax(logits, dim=1).squeeze()

        # Calculate Compound Score
        compound_score = (probabilities[1] - probabilities[0]).item()

        # Determine sentiment label
        sentiment_map = {0: 'negative', 1: 'positive'}
        predicted_class = torch.argmax(logits, dim=1).item()
        sentiment_label = sentiment_map.get(predicted_class, 'neutral')

        return sentiment_label, compound_score

# Load DataFrame
df = pd.read_csv('cleaned_singapore_airlines_reviews.csv')

# Apply the function to each review and split the results into two columns
df[['sentiment', 'compound_score']] = df['cleaned_review'].apply(
    lambda x: pd.Series(classify_sentiment_and_score(x))
)

# Save the updated DataFrame
df.to_csv('classified_reviews_with_sentiment_and_compound.csv', index=False)
print("Dataset saved as 'classified_reviews_with_sentiment_and_compound.csv'.")
```

```
Dataset saved as 'classified_reviews_with_sentiment_and_compound.csv'.
```

VADER

```python
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
import pandas as pd

# Initialize the VADER sentiment analyzer
analyzer = SentimentIntensityAnalyzer()

# Function to classify sentiment using VADER
def classify_sentiment_vader(text):
    """
    Classify sentiment and calculate compound score using VADER.

    Parameters:
        text (str): The review text to analyze.

    Returns:
        tuple: Sentiment label (positive/negative) and Compound Score.
    """
    # Get sentiment scores from VADER
    scores = analyzer.polarity_scores(text)

    # Determine sentiment based on compound score
    compound_score = scores['compound']
    if compound_score >= 0.00:
        sentiment_label = 'positive'
    elif compound_score < -0.00:
        sentiment_label = 'negative'


    return sentiment_label, compound_score

# Load the cleaned dataset
df = pd.read_csv('cleaned_singapore_airlines_reviews.csv')

# Apply the VADER sentiment analysis function to each review
df[['vader_sentiment', 'vader_compound_score']] = df['cleaned_review'].apply(
    lambda x: pd.Series(classify_sentiment_vader(x))
)

# Save the updated DataFrame with VADER results
df.to_csv('vader_sentiment_results.csv', index=False)
print("Dataset saved as 'vader_sentiment_results.csv'.")
```

```
Dataset saved as 'vader_sentiment_results.csv'.
```

Figure 5: Sentiment Analysis Framework and python coding

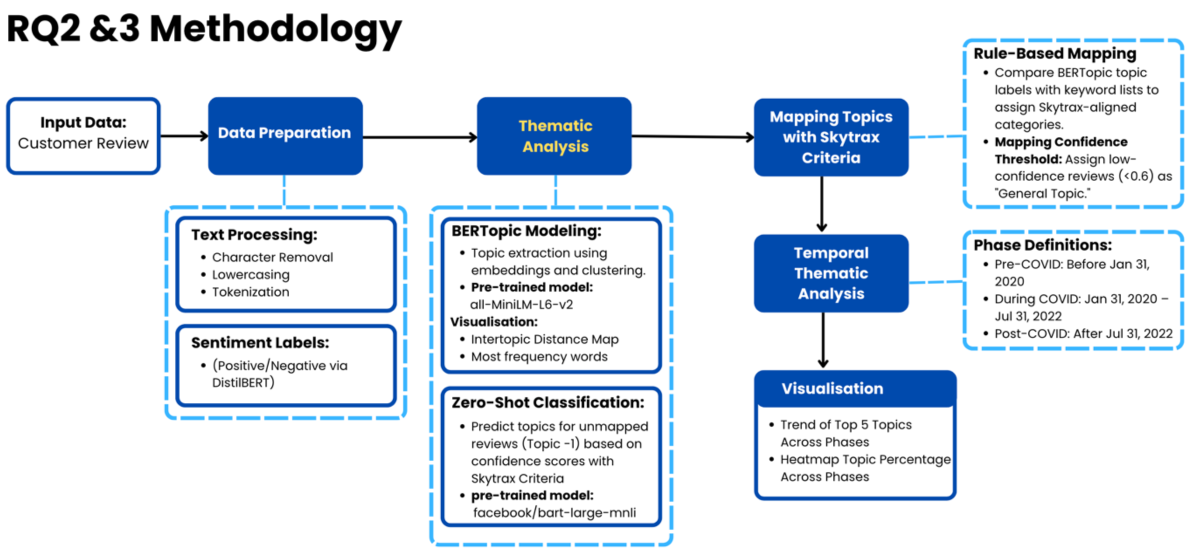
As shown in Figure 5, DistilBERT, a distilled version of BERT optimized for efficiency (Hugging Face, n.d.), and VADER were employed to analyze customer sentiment from Singapore Airlines reviews across distinct periods—Pre-COVID, During COVID-19, and Post-COVID. The model's ability to calculate compound scores, ranging from -1 to 1, captures sentiment intensity, allowing for precise classification and insights into evolving customer attitudes. Temporal analysis segmented sentiments by phases, while visualization techniques, including line charts and distribution plots, highlighted trends. DistilBERT's contextual understanding of phrases ensured accurate analysis, providing actionable insights into customer satisfaction and dissatisfaction during critical industry shifts. However, to ensure the reliability of this analysis, evaluation metrics and equations (Table 6 and 7) were employed to assess the accuracy of the model and the validity of its results.

Table 6: Confusion matrix

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Table 7: Evaluation Metrics and Formulas Used for RQ1 Sentiment Analysis

Metric	Equations	Justification
Accuracy	$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$	Measures the proportion of correctly predicted sentiment labels. Ensures an overall assessment of model performance.
Precision	$\text{Precision} = \frac{TP}{TP+FP}$	Evaluates how many of the predicted positive sentiments are correct, reducing false positives.
Recall (Sensitivity)	$\text{Recall} = \frac{TP}{TP+FN}$	Determines the ability of the model to correctly identify positive sentiments, reducing false negatives.
F1-Score	$\text{F1} = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$	Balances precision and recall, providing a harmonic mean to evaluate overall performance.
ROC-AUC Score	Area under the Receiver Operating Characteristic Curve	Measures the model's ability to distinguish between positive and negative sentiments, considering all classification thresholds.
Compound Score	Compound=P(Positive)−P(Negative)	Quantifies sentiment intensity on a scale of -1 to 1, aiding in distinguishing sentiment polarity.



1. Thematic analysis Overview

1.1 Using BERTopic

```python
import pandas as pd
from bertopic import BERTopic
from sentence_transformers import SentenceTransformer

# Load the original dataset with reviews, sentiment, and published_date
df = pd.read_csv('simplified_reviews_with_sentiment.csv')

# Ensure the 'published_date' is in datetime format
df['published_date'] = pd.to_datetime(df['published_date'], errors='coerce')
df = df.dropna(subset=['published_date'])

# Load the pre-trained model for embeddings
model = SentenceTransformer('all-MiniLM-L6-v2')

# Generate embeddings for the reviews
embeddings = model.encode(df['cleaned_review'].to_list(), show_progress_bar=True)

# Fit the BERTopic model and get topics for each review
topic_model = BERTopic()
topics, _ = topic_model.fit_transform(df['cleaned_review'].to_list(), embeddings)

# Add the topic IDs back to the original DataFrame
df['topic'] = topics

# Get topic information
topic_info = topic_model.get_topic_info()

# Map topic labels and keywords
def get_topic_details(topic_id):
    if topic_id == -1:  # Handle outliers
        return "Outlier", ""
    else:
        # Fetch topic information
        topic_row = topic_info[topic_info['Topic'] == topic_id]
        if topic_row.empty:  # If topic ID does not exist in topic_info
            return "Unknown", "N/A"
        else:
            topic_label = topic_row['Name'].values[0]
            keywords = ", ".join([kw[0] for kw in topic_model.get_topic(topic_id)])
            return topic_label, keywords

# Apply the mapping to add topic labels and keywords
df[['topic_label', 'topic_keywords']] = df['topic'].apply(
    lambda x: pd.Series(get_topic_details(x))
)

# Save the updated DataFrame with topics, labels, and keywords
df.to_csv('thematic_analysis_per_review_with_labels.csv', index=False)
print("Thematic analysis with topics, labels, and keywords saved successfully!")
```

1.2 Using Zero-Shot Classification for Topic -1

by Mapping with Skytrax Criteria

```python
from transformers import pipeline
import pandas as pd
from tqdm import tqdm
import logging

# Set up logging for errors
logging.basicConfig(filename="error_log_zero_shot.txt", level=logging.ERROR, format="%(asctime)s - %(message)s")

# Load the dataset
df = pd.read_csv('thematic_analysis_per_review_with_labels.csv')
topic_minus_one_df = df[df['topic'] == -1].reset_index(drop=True)

# Initialize Zero-Shot Classification pipeline
classifier = pipeline("zero-shot-classification", model="facebook/bart-large-mnli", multi_label=True)

# Define candidate labels
candidate_labels = [
    "Boarding assistance",
    "Service friendliness / hospitality",
    "Service attentiveness / efficiency",
    "Tickets and refunds",
    "Assisting families",
    "Online booking and check-in services",
    "Baggage delivery",
    "Seat comfort",
    "Cleanliness",
    "Meal service efficiency",
    "Entertainment",
    "Airline Lounge : product facilities"
]

# Add empty columns for results
topic_minus_one_df['predicted_topic'] = None
topic_minus_one_df['confidence'] = None

# Process each review with tqdm (Progress Bar)
for idx, review in tqdm(topic_minus_one_df['cleaned_review'].items(), desc="Classifying topics", total=len(topic_min
    try:
        # Skip empty reviews
        if not isinstance(review, str) or review.strip() == "":
            topic_minus_one_df.at[idx, 'predicted_topic'] = "General Topic"
            topic_minus_one_df.at[idx, 'confidence'] = 0.0
            continue

        # Run Zero-Shot Classification
        result = classifier(review, candidate_labels)
        topic_minus_one_df.at[idx, 'predicted_topic'] = result['labels'][0]
        topic_minus_one_df.at[idx, 'confidence'] = result['scores'][0]

    except Exception as e:
        # Log error and assign default values
        logging.error(f"Error processing index {idx}: {e}")
        topic_minus_one_df.at[idx, 'predicted_topic'] = "General Topic"
        topic_minus_one_df.at[idx, 'confidence'] = 0.0

# Save the fixed dataset
topic_minus_one_df.to_csv('zero_shot_classified_topic_minus_one_fixed.csv', index=False)
print("\nClassified topics saved successfully as 'zero_shot_classified_topic_minus_one_fixed.csv'")
```

```
Classifying topics: 100%|██████████████████| 5696/5696 [1:14:40<00:00,  1.27it/s]
```

Figure 6: Thematic Analysis Framework and python coding

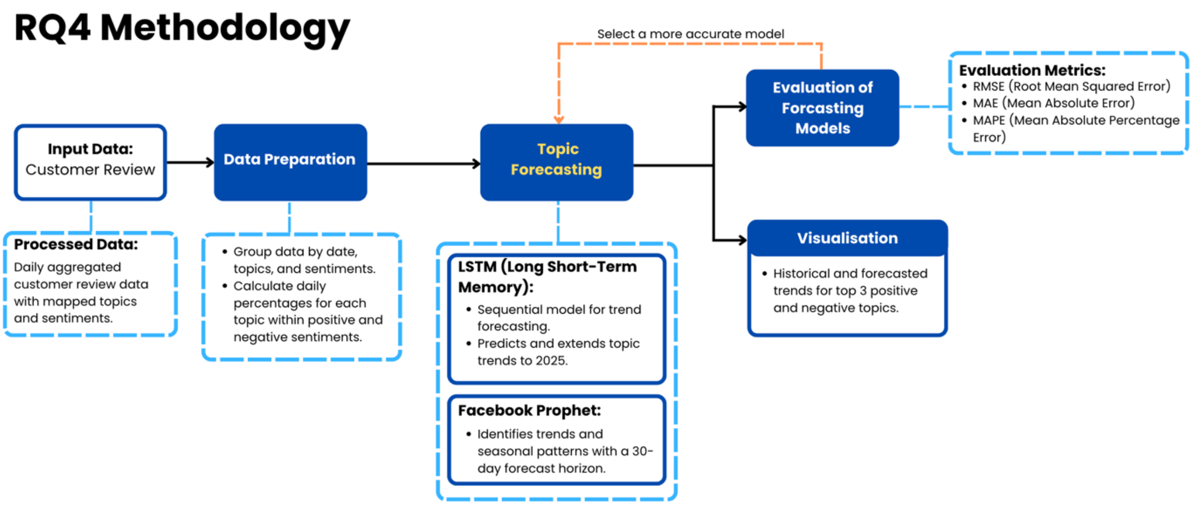
The framework in Figure 6 combines advanced NLP techniques to analyze customer sentiments and thematic insights from reviews, aligning with Skytrax criteria to address RQ2 and RQ3 effectively.

BERTopic Modeling leverages the “all-MiniLM-L6-v2” embedding model to extract themes such as “Service Efficiency” and “Cleanliness.” This unsupervised method clusters reviews based on their content without predefined criteria, offering flexibility and enabling the discovery of diverse topics. Visual tools like intertopic distance maps and keyword frequencies provide further understanding of these clusters.

Zero-Shot Classification, using “facebook/bart-large-mnli,” categorizes outlier reviews (topic -1) to enhance thematic coverage. This approach leverages Zero-Shot Learning, enabling the model to assign predefined topics without requiring domain-specific training, making it particularly useful for categorizing unseen or ambiguous data. Its Multi-Label Classification capability allows the identification of multiple relevant themes in a single review, capturing the complexity of customer feedback. Additionally, the model’s advanced Context Understanding ensures accurate topic assignment by analyzing the semantic relationships between the review text and candidate labels. Reviews with confidence scores below 0.6 are labeled as “General Topic” to ensure data reliability. Despite its high accuracy, Zero-Shot Classification is computationally intensive, making it impractical for large datasets on its own. By combining this technique with BERTopic, a balance between thematic discovery and precision mapping is achieved, optimizing resource usage and enhancing the overall analytical depth.

Temporal Thematic Analysis segments reviews into Pre-COVID, During COVID, and Post-COVID phases, revealing shifts in customer sentiment and recurring themes over time. Visualizations such as heatmaps and trend charts highlight the top topics’ evolution, offering actionable insights.

This hybrid approach balances thematic exploration and classification accuracy while addressing computational limitations, ensuring a comprehensive analysis of customer perceptions to guide service enhancements.



1.1 Training with LSTM (Long Short-Term Memory)

```python
import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense
import matplotlib.pyplot as plt

# Load the preprocessed data
file_path = 'daily_topic_sentiment_percentages.csv'  # Replace with your file path
df = pd.read_csv(file_path)

# Function to prepare data and train LSTM for each topic
def train_lstm_for_topic(topic, sentiment, sequence_length=30):
    # Filter data for a specific topic and sentiment
    filtered_df = df[(df['mapped_topic'] == topic) & (df['sentiment'] == sentiment)]

    # Check if there is enough data for training
    if len(filtered_df) <= sequence_length:
        print(f"Not enough data for Topic: {topic}, Sentiment: {sentiment}")
        return None, None, None

    # Ensure data is sorted by date
    filtered_df['published_date'] = pd.to_datetime(filtered_df['published_date'])
    filtered_df = filtered_df.sort_values('published_date')

    # Extract percentage values
    percentage_values = filtered_df['Percentage'].values.reshape(-1, 1)

    # Normalize data
    scaler = MinMaxScaler()
    percentage_scaled = scaler.fit_transform(percentage_values)

    # Create sequences
    X, y = [], []
    for i in range(len(percentage_scaled) - sequence_length):
        X.append(percentage_scaled[i:i+sequence_length])
        y.append(percentage_scaled[i+sequence_length])
    X, y = np.array(X), np.array(y)

    # Build the LSTM model
    model = Sequential([
        LSTM(50, return_sequences=True, input_shape=(X.shape[1], X.shape[2])),
        LSTM(50),
        Dense(1)
    ])
    model.compile(optimizer='adam', loss='mse')

    # Train the model
    model.fit(X, y, epochs=20, batch_size=32, verbose=0)

    # Predict using the model
    predictions = model.predict(X)
```

```python
    # Reverse normalization for predictions and actual values
    predicted_values = scaler.inverse_transform(predictions)
    actual_values = scaler.inverse_transform(y.reshape(-1, 1))

    return filtered_df.iloc[sequence_length:], actual_values, predicted_values

# Define top 3 topics for each sentiment
top_positive_topics = [
    'Service attentiveness / efficiency',
    'Meal service efficiency',
    'Seat comfort'
]
top_negative_topics = [
    'Service attentiveness / efficiency',
    'Tickets and refunds',
    'Seat comfort'
]

# Function to plot results for multiple topics
def plot_results(topics, sentiment, colors):
    plt.figure(figsize=(14, 7))
    for idx, topic in enumerate(topics):
        # Train and forecast for each topic
        dates, actual_values, predicted_values = train_lstm_for_topic(topic, sentiment)

        if dates is None:  # Skip topics with insufficient data
            continue

        # Plot actual values
        plt.plot(dates['published_date'], actual_values, linestyle='--', label=f'{topic} - Actual',
                 color=colors[idx], alpha=0.3, linewidth=1)

        # Plot predicted values
        plt.plot(dates['published_date'], predicted_values, label=f'{topic} - Predicted',
                 color=colors[idx], alpha=1, linewidth=2)

    # Add labels and legend
    plt.title(f'LSTM Forecasting for Top 3 Topics ({sentiment.capitalize()})')
    plt.xlabel('Date')
    plt.ylabel('Percentage')
    plt.legend(loc='best', title='Topics')
    plt.grid()
    plt.show()

# Define colors for each topic
colors = ['blue', 'orange', 'green']

# Plot for positive sentiment
plot_results(top_positive_topics, 'positive', colors)

# Plot for negative sentiment
plot_results(top_negative_topics, 'negative', colors)
```

1.2 Training with Facebook Prophet

```python
import numpy as np
import pandas as pd
from prophet import Prophet
import matplotlib.pyplot as plt
from sklearn.metrics import mean_squared_error, mean_absolute_error

# Load the preprocessed data
file_path = 'daily_topic_sentiment_percentages.csv'  # Replace with your file path
df = pd.read_csv(file_path)

# Function to prepare data and train Prophet for each topic
def train_prophet_for_topic(topic, sentiment):
    # Filter data for a specific topic and sentiment
    filtered_df = df[(df['mapped_topic'] == topic) & (df['sentiment'] == sentiment)]

    # Check if there is enough data for training
    if len(filtered_df) <= 10:  # Ensure sufficient data
        print(f"Not enough data for Topic: {topic}, Sentiment: {sentiment}")
        return None, None, None

    # Prepare data for Prophet
    filtered_df['published_date'] = pd.to_datetime(filtered_df['published_date'])
    filtered_df = filtered_df.sort_values('published_date')
    prophet_df = filtered_df[['published_date', 'Percentage']].rename(
        columns={'published_date': 'ds', 'Percentage': 'y'}
    )

    # Train the Prophet model
    model = Prophet()
    model.fit(prophet_df)

    # Create a dataframe for future predictions
    future = model.make_future_dataframe(periods=30)  # Forecast next 30 days
    forecast = model.predict(future)

    # Match predictions with actual values
    actual_values = prophet_df['y']
    predicted_values = forecast.loc[:len(actual_values)-1, 'yhat']  # Match with histor

    return prophet_df, actual_values, predicted_values

# Define top 3 topics for each sentiment
top_positive_topics = [
    'Service attentiveness / efficiency',
    'Meal service efficiency',
    'Seat comfort'
]
top_negative_topics = [
    'Service attentiveness / efficiency',
    'Tickets and refunds',
    'Seat comfort'
]
```

```python
# Function to plot results for multiple topics
def plot_results(topics, sentiment, colors):
    plt.figure(figsize=(14, 7))
    for idx, topic in enumerate(topics):
        # Train and forecast for each topic
        data, actual_values, predicted_values = train_prophet_for_topic(topic, sentiment)

        if data is None:  # Skip topics with insufficient data
            continue

        # Plot actual values
        plt.plot(data['ds'], actual_values, linestyle='--', label=f'{topic} - Actual',
                 color=colors[idx], alpha=0.3, linewidth=1)

        # Plot predicted values
        plt.plot(data['ds'], predicted_values, label=f'{topic} - Predicted',
                 color=colors[idx], alpha=1, linewidth=2)

    # Add labels and legend
    plt.title(f'Prophet Forecasting for Top 3 Topics ({sentiment.capitalize()})')
    plt.xlabel('Date')
    plt.ylabel('Percentage')
    plt.legend(loc='best', title='Topics')
    plt.grid()
    plt.show()

# Define colors for each topic
colors = ['blue', 'orange', 'green']

# Plot for positive sentiment
plot_results(top_positive_topics, 'positive', colors)

# Plot for negative sentiment
plot_results(top_negative_topics, 'negative', colors)
```

Figure 7: Topic Forcasting Framework and python coding

Table 8: Evaluation Metrics and Formulas Used for RQ4 Forecasting Models

Metric	Equations	Justification
RMSE	$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$	Measures the average magnitude of prediction errors; lower RMSE indicates better model accuracy.
MAE	$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\lvert\hat{y}_i - y_i\rvert$	Measures the average absolute error; lower MAE indicates higher accuracy.
MAPE	$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{\hat{y}_i - y_i}{y_i}\right\rvert \times 100$	Measures error as a percentage; lower MAPE indicates better relative accuracy.

As outlined in Figure 7 and Table 8, this study employs LSTM and Facebook Prophet models to forecast the prevalence of main topics associated with customer sentiments, with the goal of identifying the more accurate forecasting approach. Both models were evaluated using RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error). These metrics provide a quantitative assessment of prediction accuracy, with lower values indicating better performance. Specifically, smaller RMSE and MAE values reflect fewer overall errors, while lower MAPE demonstrates higher prediction precision relative to actual values.

To ensure comparability across varying time periods, the study employs Prediction Scaling, normalizing topic counts by day and sentiment. This is expressed mathematically as:

$$\text{Percentage} = \frac{\text{Topic Count}}{\text{Total Daily Count}} \times 100$$

By transforming raw topic counts into percentages, this approach ensures that trends focus on relative changes in topic prevalence, rather than absolute variations influenced by fluctuating review volumes. Both LSTM and Facebook Prophet leverage this scaled data for consistent trend analysis.

LSTM excels in analyzing dynamic datasets with complex, non-linear dependencies and long-term patterns, making it ideal for predicting granular changes, such as shifts in sentiments related to 'Service Attentiveness/Efficiency.' However, it requires significant computational resources and longer training times. Conversely, Facebook Prophet is optimized for time-series data with clear seasonality and irregular patterns, such as trends in 'Tickets and Refunds' during holiday seasons or travel disruptions. Its interpretable outputs and efficiency in handling periodic patterns make it suitable for identifying recurring refund peaks during crises like COVID-19. However, it struggles with capturing intricate, non-linear dependencies.

The evaluation compares both models' performance on key topics to determine the superior method. Preliminary results suggest that LSTM’s ability to learn intricate relationships often yields more precise predictions, while Prophet excels in capturing general trends. The rigorous evaluation ensures that the selected model will guide Singapore Airlines in anticipating future feedback trends, prioritizing service improvements, and maintaining its competitive position. By employing these methods and metrics, this study delivers actionable insights for strategic decision-making.

Dataset Description

The "Singapore Airlines Reviews" dataset aggregates 10,000 anonymized TripAdvisor reviews, offering insights into customer experiences. It was sourced from Kaggle, published by user Kanchana1990 on 13 March 2024. This dataset contains critical review attributes detailed in the table 9.

Table 9: Data dictionary of reviews

Column Name	Description
published_date	Date and time of review publication.
published_platform	Platform where the review was posted.
rating	Customer satisfaction rating (1 to 5).
type	Specifies the content as a review.
text	Detailed customer feedback.
title	Summary of the review.
helpful_votes	Number of users finding the review helpful.

Representativeness of the Dataset

The dataset captures a globally diverse range of perspectives, with contributions from users in Asia, North America, Europe, and Australia. While TripAdvisor enables users to provide demographic details such as age, gender, and geographic location, this dataset employs PII anonymization to protect user privacy. Consequently, such demographic attributes are unavailable. However, this does not compromise the dataset's representativeness of Singapore Airlines' international customer base, as TripAdvisor is a widely used platform that reflects a culturally diverse audience (Vora, 2023). The reviews are exclusively in English, which ensures consistency and aligns with the capabilities of the sentiment analysis models employed in this study, such as DistilBERT. Although this choice may limit linguistic diversity, it allows for robust and reliable text analysis while addressing a significant portion of TripAdvisor's global user base.

The data underwent thorough preparation to ensure quality and consistency. Only one column, "title," had a single missing value, which does not impact the analysis as the primary focus is on the "text" column, containing the review content. The text preprocessing included removing HTML tags, special characters, and stopwords, converting all text to lowercase for consistency, and tokenizing the text using NLTK tools. Furthermore, normalization steps, such as spelling correction and text standardization, were applied to address inconsistencies. This rigorous data cleansing process ensured the dataset's readiness for sentiment and thematic analysis.

Justification for TripAdvisor as the Data Source

TripAdvisor was chosen as the data source due to its dominant role in the travel and tourism industry, where it is recognized for verified, first-hand customer reviews. Unlike platforms such as Google Reviews, which primarily focus on business reviews and lack extensive airline-specific data, TripAdvisor specializes in detailed feedback from travelers. Google Flights, while partnering with over 300 airlines, focuses on price comparisons rather than providing a platform for customer reviews (Google, 2025). Additionally, Yelp, though popular in North America and parts of Europe, has minimal traction in Asia (Lauchlan, 2016), reducing its relevance for studying airlines like Singapore Airlines, which operates globally and has a main hub in Asia. TripAdvisor, on the other hand, ranks as one of the top 10 travel intermediaries in Asia-Pacific by revenue and is the only platform among these to offer extensive airline review data (GlobalData, n.d.; Vora, 2023).

The inclusion of TripAdvisor data ensures access to high-quality, diverse reviews. Its popularity ranking algorithm, based on the quality, quantity, and recency of reviews, further enhances the reliability of this dataset for sentiment and thematic analysis (Vora, 2023).

Data Quality and Preprocessing

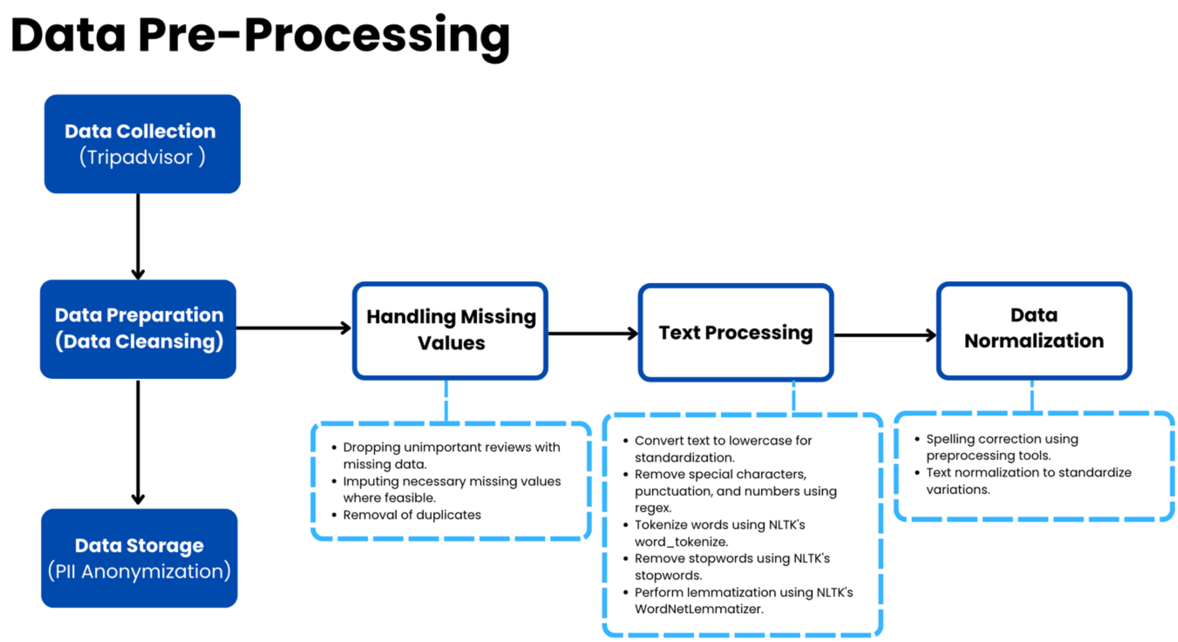


Figure 8: Data pre-processing for sentiment analysis

According to Figure 8, the dataset underwent rigorous preprocessing to ensure reliability for analysis. Duplicate entries were removed, and text data was cleansed through tokenization, removal of special characters, and stopword elimination. Lemmatization was applied to standardize words, improving machine learning compatibility. These steps addressed common issues in user-generated content, such as inconsistencies and irrelevant elements, resulting in a clean, normalized dataset. The preprocessed data was saved for further analysis, ensuring textual uniformity and accuracy in sentiment extraction.

Results Analysis and Discussion

This section delineates the study's findings, which sought to analyse customer sentiments articulated in SIA reviews based on their experiences. The results are structured in accordance with the research questions presented in the introduction.

Exploratory Data Analysis (EDA) of Dataset

```
Dataset contains 10000 rows and 7 columns.
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 7 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   published_date      10000 non-null  object
 1   published_platform  10000 non-null  object
 2   rating              10000 non-null  int64
 3   type                10000 non-null  object
 4   text                10000 non-null  object
 5   title               9999 non-null   object
 6   helpful_votes       10000 non-null  int64
dtypes: int64(2), object(5)
memory usage: 547.0+ KB
Missing values per column:
published_date        0
published_platform    0
rating                0
type                  0
text                  0
title                 1
helpful_votes         0
dtype: int64
```

Figure 9: Summary statistics of dataset

Figure 9 shows that the dataset contains 10,000 rows and 7 columns, with no missing values except for one null entry in the title column. Key attributes include review text, customer rating, and published_date. The dataset's completeness supports robust analysis and minimizes preprocessing needs.

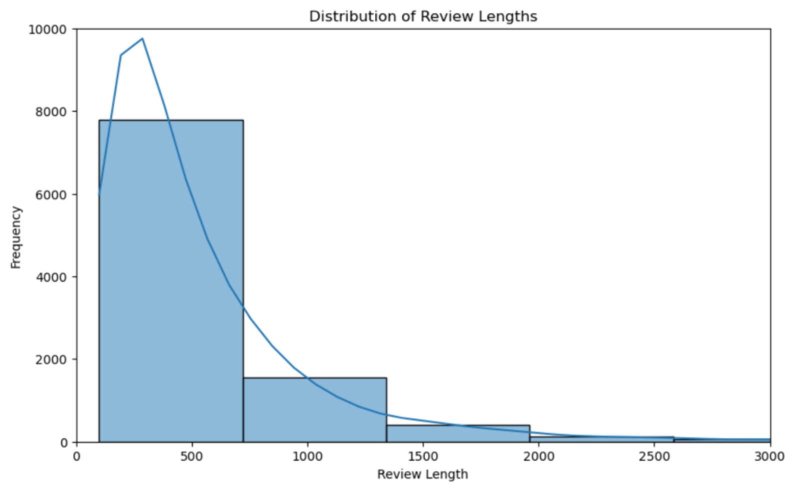


Figure 10: Distribution of Review Lengths

Figure 10 illustrates the frequency of review lengths, revealing a right-skewed distribution where most reviews are under 500 characters. The superimposed KDE curve highlights a smooth probability density, emphasizing the concentration of shorter reviews and the gradual decline in frequency for longer reviews. This pattern indicates that customers tend to provide concise feedback, with fewer detailed reviews extending beyond 1,000 characters.

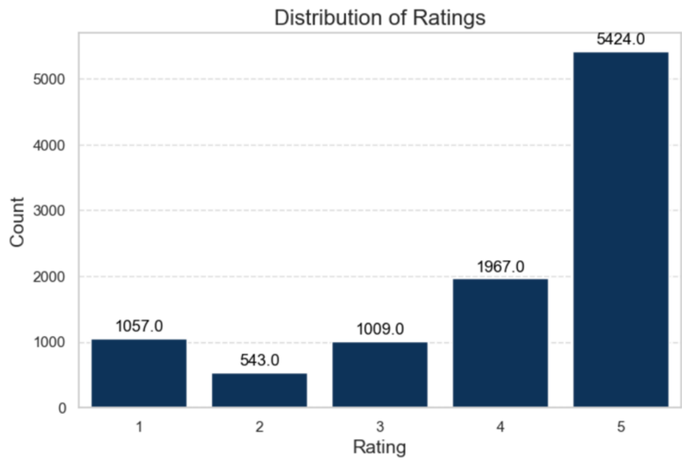


Figure 11: Distribution of Ratings

Figure 11 shows that 5-star ratings dominate (54.2%), reflecting high customer satisfaction. Lower ratings (1-2 stars) constitute a minority, emphasizing areas of improvement in specific service aspects. This aligns with Singapore Airlines' reputation as a top-tier airline but highlights potential for addressing negative feedback.

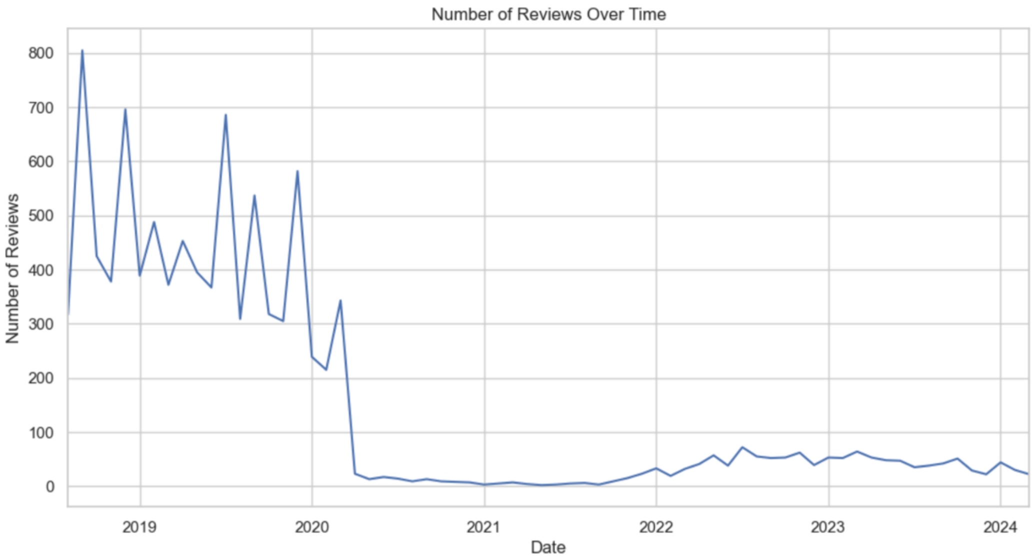


Figure 12: Temporal Distribution of Reviews

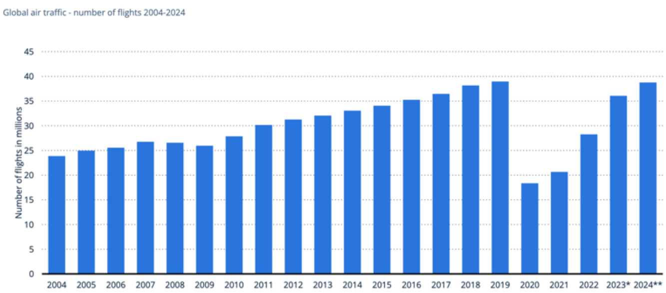


Figure13: Number of flight performed by the global airline (Statista, 2024)

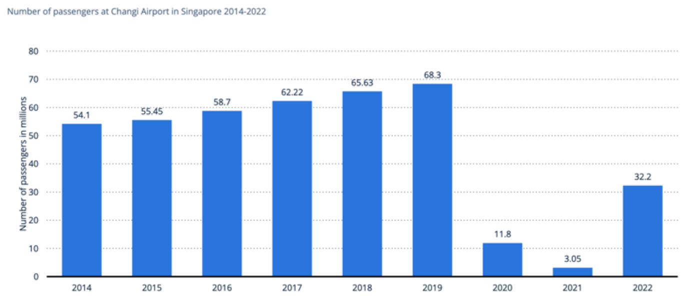


Figure14: Number of passenger at Changi Airport in Singapore (Statista, 2024)

The sharp decline in reviews during 2020-2021, as seen in Figure 12, corresponds to the global travel restrictions during the COVID-19 pandemic. This trend aligns with the significant reduction in global flight operations, as depicted in Figure 13, where the number of flights dropped sharply from 38.9 million in 2019 to 16.4 million in 2020. At Changi Airport, Singapore's main hub for Singapore Airlines, passenger numbers dropped to only 17.28% of the previous year as shown in Figure 14. In May 2020, Singapore Airlines reported (see Appendix A) a 99.6% year-on-year decline in passenger carriage and a passenger load factor of only 8.6% (Singapore Airlines, 2020). Although flights resumed post-COVID in August 2022, passenger numbers remained less than half of pre-pandemic levels, leading to a continued reduction in reviews.

Figure 15: Commonly Mentioned Words in Reviews

The word cloud (Figure 15) highlights key topics such as “service,” “flight,” and “crew,” indicating customer emphasis on in-flight experience and staff performance. Terms like “food” and “seat” suggest that amenities and comfort remain critical aspects of customer evaluations. However, frequently mentioned terms may reflect both positive and negative experiences. Therefore, incorporating sentiment analysis will help identify specific strengths and weaknesses associated with these topics, providing a more nuanced understanding of customer perceptions and actionable insights for the airline.

Sentiment Analysis Results Across Time Periods (RQ1)

Model Selection for Sentiment Analysis

Table 10 : Comparison between DistilBERT and VADER Sentiment Distribution

Sentiment	DistilBERT		VADER	
	Counts	Percentages	Counts	Percentages
Positive	5625	56.25%	8679	86.79%
Negative	4375	43.74%	1321	13.21%

The sentiment classification results are presented in Table 10. DistilBERT identified 56.25% of reviews as positive and 43.75% as negative, offering a balanced sentiment distribution. In contrast, VADER classified 86.79% of reviews as positive, potentially oversimplifying customer feedback trends. To assess the accuracy of both models, mismatched sentiments were analyzed and compared, as shown in Table 11.

Table 11: Random mismatch sentiment analysis between two analytics tools

User	Review text	Sentiment Model	
		DistilBERT	VADER
1	As a wheelchair person made to wait till last to board and so no overhead space for carryon bags. Airbus 350 SIN-BOM coach seats are too close and can't lower the trays. Limited toilets long wait for toilets.	negative	positive
2	Premium is not premium no leg room cramped and crowded uncomfortable seats and no better than economy which is cheaper not worth the extra money.	negative	positive
3	I have been trying to book an award ticket with KrisFlyer miles. Impossible! waitlist after waitlist. Unacceptable! Either redeem almost double the required miles or pay! Not good Singapore Airlines.	negative	positive
4	Singapores airline service has certainly dropped and dealing with them is very hard	negative	positive
5	I used to be a fan of singapores but no more	negative	positive
6	Good flight, service was was excellent & the staff were the best I have encountered. Food wad good without being exceptional. Only complaint was inhouse entertainment system which was very cumbersome to use & selection was limited.	negative	positive
7	We where a bit disapointed with the service and food on this flight. Usually we are very impressed with them but this time they let us down.	negative	positive
8	The experience on Singapore airlines was second to none . The staff were excellent and courteous . Would highly recommend	positive	negative
9	This is first time I took Singapore Airline, which was well organized. For boarding the airline divided people into several groups. This avoided crowded stiuation.	positive	negative
10	From the time you get into the business class lounge until you board the plane the service is wonderful...the food choice is 5 star and the service is also 5 star but you get what you pay for	positive	negative

Table 11 highlights ten examples, randomly selected, where the two models disagreed on sentiment classification. Notably, VADER often classified critical reviews as positive, while DistilBERT provided sentiment classifications more aligned with the textual content. For instance, in reviews containing both positive and negative aspects (e.g., user 6), DistilBERT identified the sentiment as negative, which better reflects areas needing improvement. This capability makes DistilBERT more suitable for deriving actionable insights to enhance service quality.

```
DistilBERT Evaluation (Surrogate Model):
              precision    recall  f1-score   support

           0       0.97      0.95      0.96      4275
           1       0.96      0.98      0.97      5601

    accuracy                           0.96      9876
   macro avg       0.96      0.96      0.96      9876
weighted avg       0.96      0.96      0.96      9876

ROC-AUC Score: 0.9941629888564176

VADER Evaluation (Surrogate Model):
              precision    recall  f1-score   support

         0.0       0.29      0.95      0.45      1283
         1.0       0.99      0.66      0.79      8593

    accuracy                           0.69      9876
   macro avg       0.64      0.80      0.62      9876
weighted avg       0.90      0.69      0.74      9876

ROC-AUC Score: 0.8732009568592464

Summary Comparison:
DistilBERT ROC-AUC: 0.9941629888564176
VADER ROC-AUC: 0.8732009568592464
```

Figure 16: Comparison Confusion Metrics

To further evaluate the suitability of the models for this study, Confusion Metrics were tested, as shown in Figure 16. DistilBERT demonstrated superior performance with an accuracy of 96% and a higher ROC-AUC score of 0.99, compared to VADER’s accuracy of 69% and ROC-AUC score of 0.87. These results highlight DistilBERT's ability to handle nuanced review text, making it more appropriate for sentiment classification in this analysis. However, it is important to note that DistilBERT’s accuracy is not 100%, as the surrogate model approach may inherit limitations from the training data and does not represent absolute ground truth. Despite this, its balanced performance metrics make it a reliable choice for sentiment analysis in this context.

In conclusion, DistilBERT was selected for its higher accuracy, nuanced performance, and balanced sentiment distribution, ensuring reliable insights into customer feedback trends across different time periods.

Exploratory Data Analysis (EDA) of Sentiment Analysis

Sentiment Distribution

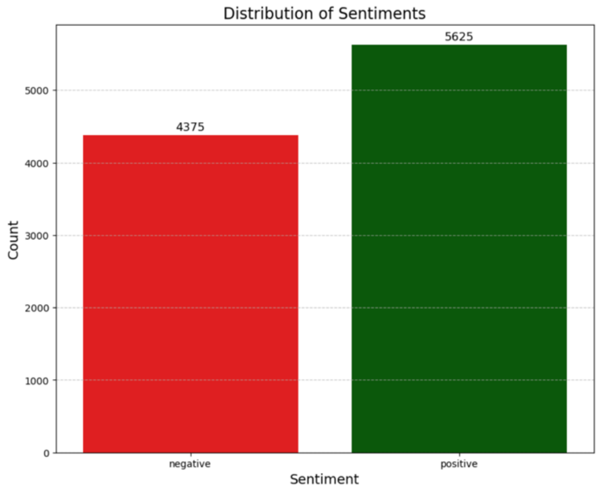


Figure 17: Sentiment Analysis using DistilBERT

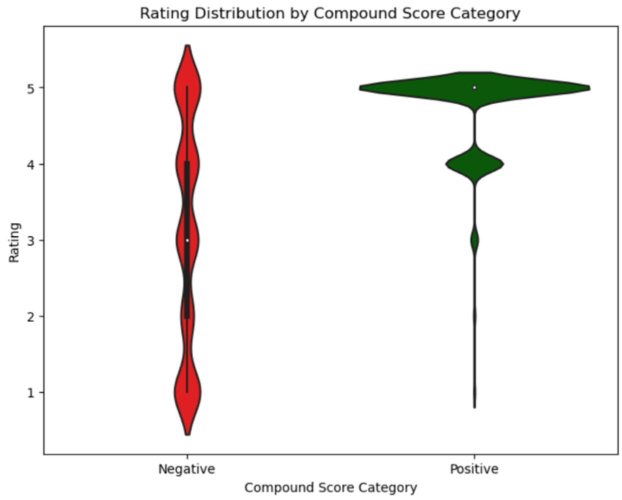


Figure 18: Rating Distribution

The sentiment analysis using DistilBERT identified 56.25% of reviews as positive and 43.75% as negative, as shown in Figure 17. This distribution highlights customer satisfaction with some service aspects while pointing to areas for improvement, providing a clear basis for identifying strengths and weaknesses.

Further insights are gained by examining the relationship between sentiment and ratings, as illustrated in Figure 18. The violin plot highlights that positive sentiments strongly correlate with higher ratings (primarily 4 and 5 stars), while negative sentiments appear across all ratings, including higher ratings (4 and 5 stars). This reflects findings from the model selection phase, where DistilBERT demonstrated the ability to interpret reviews containing both positive and negative feedback as negative when critical aspects were highlighted. This explains the presence of negative sentiments even in reviews with overall high ratings, as customers often express satisfaction while still noting areas for improvement. This nuanced interpretation reinforces the model's ability to capture complex customer feedback.

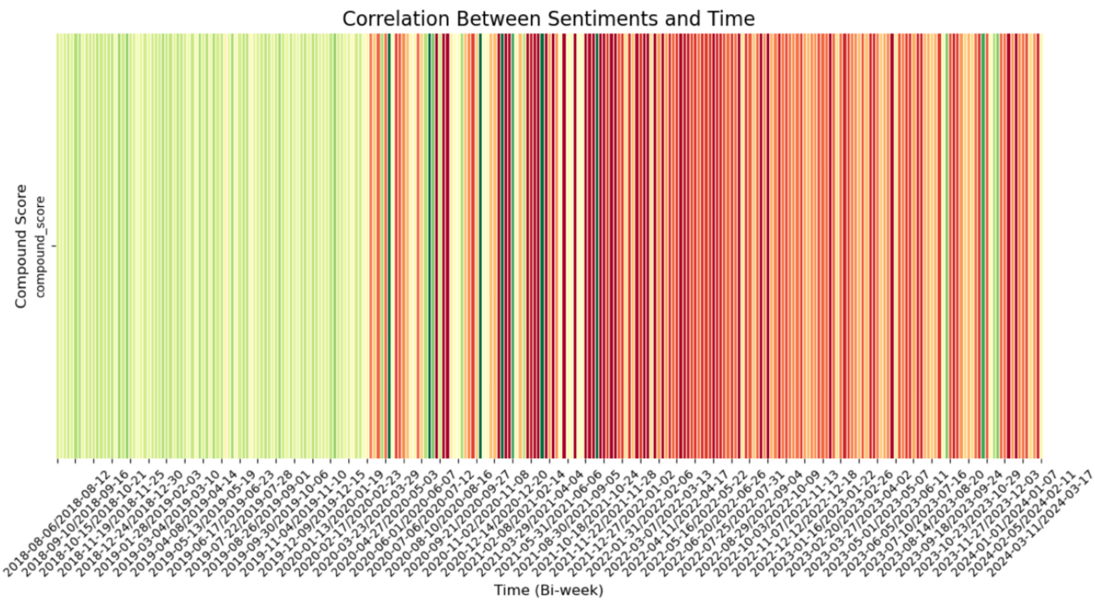


Figure 19: Correlation Between Sentiments across Time

Figure 19 illustrates the correlation between sentiments (measured by compound scores) and time, showing a sharp increase in negative sentiments immediately following the outbreak of COVID-19. This trend reflects the findings of Sobieralski (2020), who identified widespread disruptions and capacity reductions in the aviation industry during the pandemic. Pereira et al. (2023) also emphasized how crises like COVID-19 significantly influence customer sentiment, with negative emotions becoming more prominent as the industry faced operational and economic challenges.

Sentiment Distribution by Phase

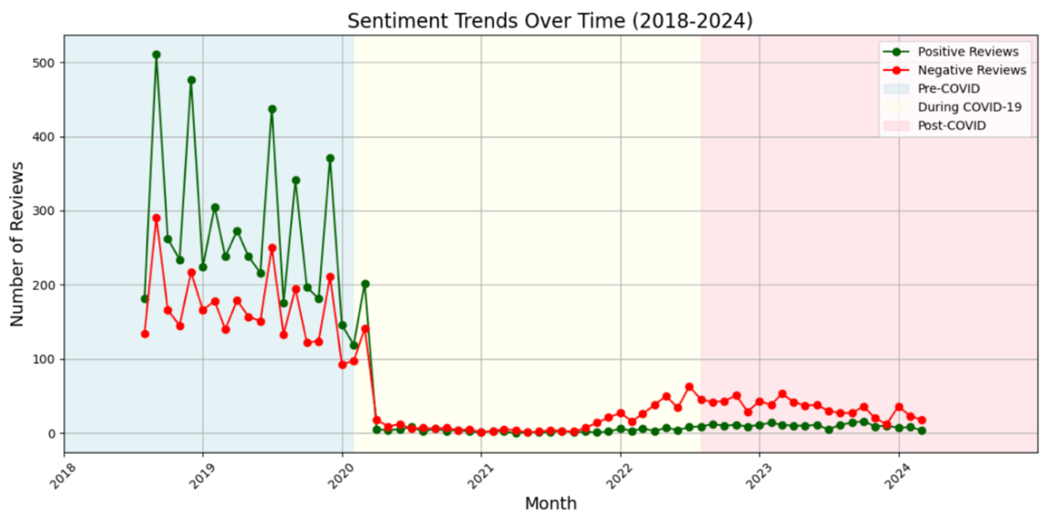


Figure 20: Sentiment Trends Over Time

Figure 20 explores sentiment trends across different time periods, categorizing reviews into pre-COVID, during COVID-19, and post-COVID phases. A significant drop in both positive and negative reviews during the pandemic (2020-2022) aligns with global travel restrictions and reduced flight operations.

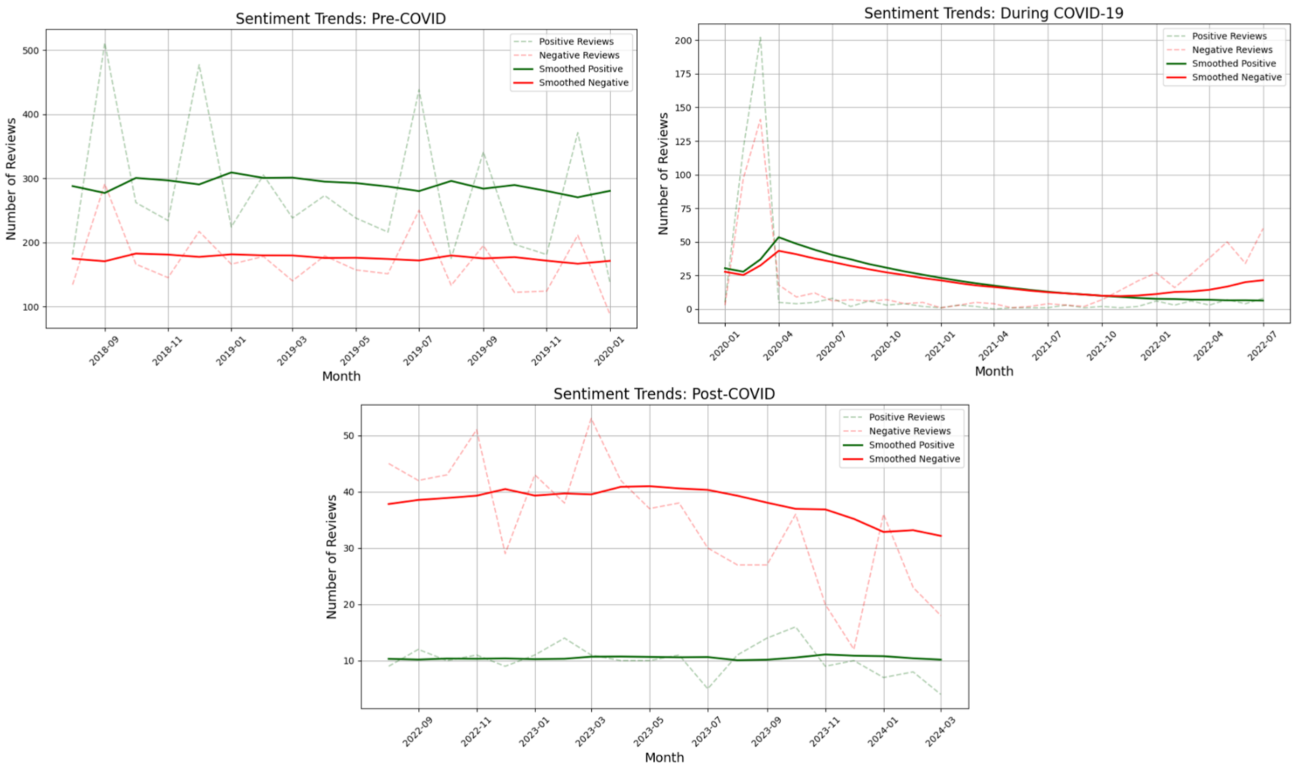


Figure 21: Shifting Sentiment Trends in each Phase

Figure 21 illustrates shifting sentiment trends over time, highlighted by the EST smoothed curves. During the pre-COVID period, the majority of customer feedback was positive, reflecting satisfaction with Singapore Airlines' services. However, the onset of COVID-19 marked a significant shift, with negative sentiments surpassing positive ones due to disruptions in travel and operational challenges. This trend continued into the post-COVID phase, where negative feedback remained dominant despite improvements in travel activity.

Figure 22: Top frequency words

To further analyze changing customer sentiments across phases, word clouds in Figure 22 illustrate the most mentioned keywords during pre-COVID, during COVID-19, and post-COVID periods. In the pre-COVID phase, positive feedback centered on "food," "comfortable," and "crew," reflecting satisfaction with service and comfort. Negative keywords like "economy" and "business" highlighted areas of concern. During COVID-19, keywords such as "ticket" and "refund" emerged in negative reviews, reflecting frustrations with cancellations and refund policies. This aligns with Pereira (2023) and Dada et al. (2021), who noted that airlines intentionally hinder refund processes by offering vouchers or delaying responses, intensifying customer dissatisfaction. Post-COVID, while positive feedback returned to themes like "food" and "crew," negative reviews continued to emphasize "ticket" and introduced "luggage" as a new area of dissatisfaction, indicating ongoing operational challenges and service recovery issues in the post-pandemic aviation sector.

In conclusion, sentiment analysis revealed evolving customer perceptions across pre-COVID, during COVID-19, and post-COVID periods, highlighting key areas of satisfaction and dissatisfaction. Building on these insights, the next section applies thematic analysis with topic modeling to uncover specific themes and topics driving customer feedback, enabling deeper understanding and actionable recommendations.

Thematic Analysis: Key Themes from SIA Reviews (RQ2 &3)

Thematic analysis identifies key topics in Singapore Airlines reviews, uncovering satisfaction drivers and aligning them with Skytrax Criteria. Using BERTopic and Zero-Shot Classification, it examines themes across pre-COVID, during COVID-19, and post-COVID phases, offering actionable insights to enhance service quality and maintain industry leadership.

Key Themes Using BERTopic

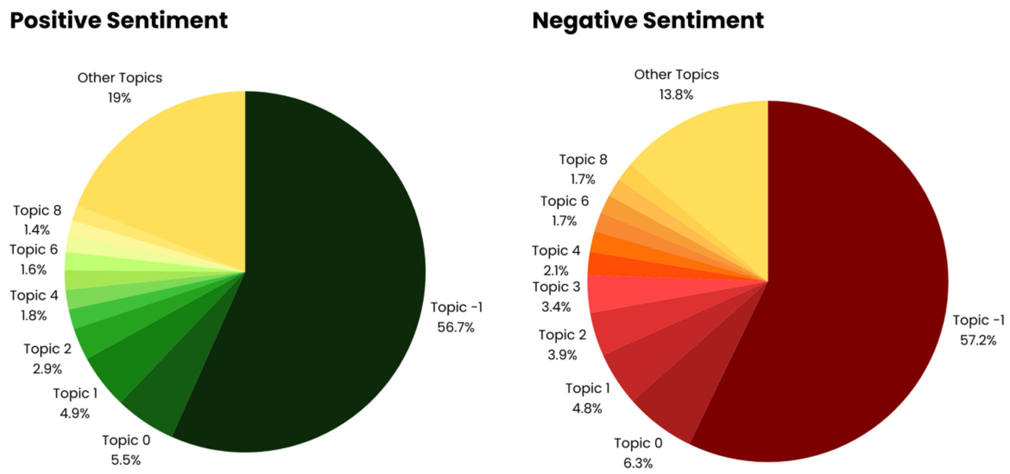


Figure 23: Proportion of Reviews for each topic

Thematic analysis using BERTopic reveals key insights into Singapore Airlines reviews, highlighting the distribution of topics across sentiments. Figure 23 illustrates the proportional breakdown of positive and negative sentiments across topics. Notably, Topic -1 accounts for over half of all reviews, reflecting the inherent challenge of categorizing reviews containing multiple or ambiguous themes. While BERTopic excels in precision and granularity, its limitations in addressing multi-faceted reviews underline the necessity of complementary methodologies.

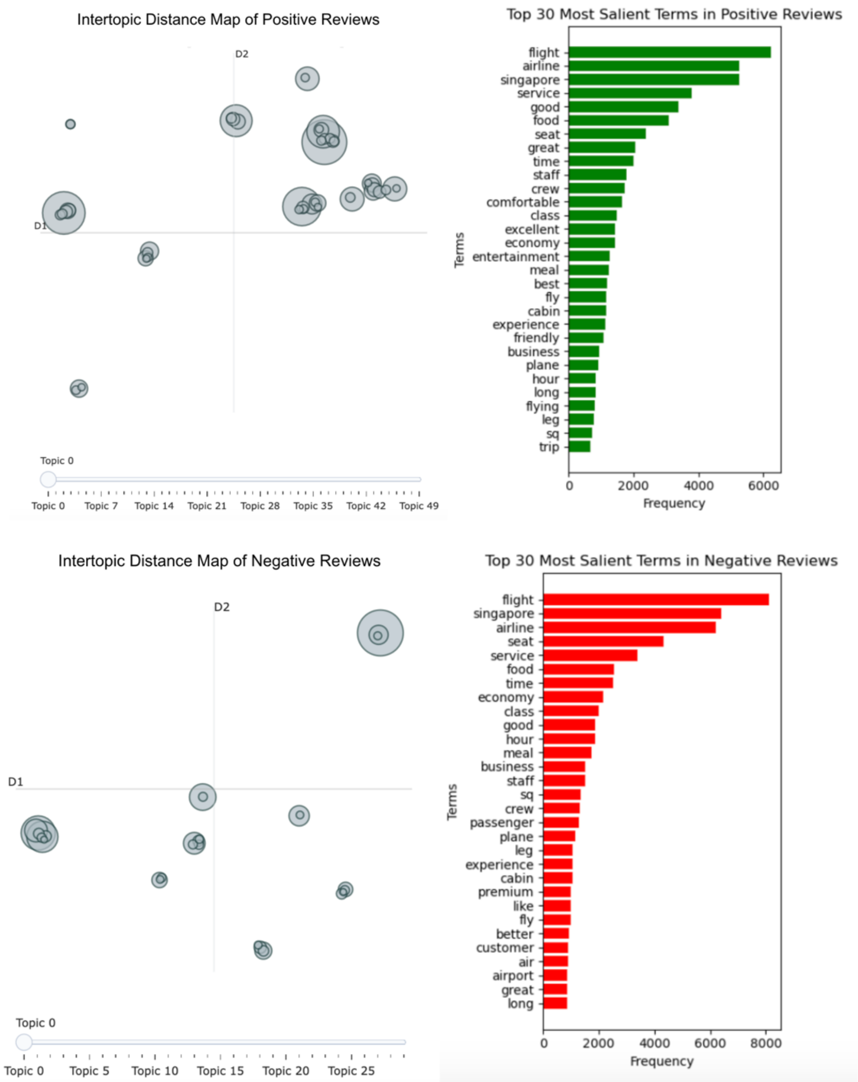


Figure 24: Intertopic distance map and Top 30 most salient terms

The Intertopic Distance Map offers a visual representation of the diversity and clustering of topics identified in positive and negative reviews. BERTopic's advanced embedding techniques enable nuanced differentiation between closely related topics, providing valuable insights into overlapping themes. However, the map also reveals several clusters positioned closely together, signifying shared attributes or thematic overlaps among topics. For instance, in-flight service and cabin comfort may share similar descriptors, whether in the form of commendations in positive reviews or constructive feedback in negative reviews. This overlap highlights the interconnected nature of certain customer experiences.

Table 12: Top 10 topics for positive reviews and key findings

Top 10 Topics for Positive Reviews			
Topic	**Topic label**	**Keywords**	**Key findings**
0	0_food_singapore_airline_service	food, singapore, airline, service, good, excellent, great, staff, fly, friendly	The reviews highlight exceptional service and high-quality food provided by Singapore Airlines. The attentive and friendly cabin crew, alongside well-prepared meals catering to diverse preferences, significantly contribute to a superior travel experience, fostering strong customer loyalty.
1	1_airline_singapore_always_service	airline, singapore, always, service, fly, best, flown, flying, experience, air	This topic highlights the consistent excellence of Singapore Airlines' service, emphasizing smooth operations, attentive staff, and reliable quality. Customers frequently commend the airline for its exceptional service standards, both in-flight and on the ground, fostering loyalty and satisfaction.
2	2_sydney_singapore_melbourne_brisbane	sydney, singapore, melbourne, brisbane, london, australia, flight, via, perth, flew	This topic highlights Singapore Airlines' exceptional service across flights connecting Australian cities to international destinations via Singapore. Customers frequently commend the airline's efficient operations, attentive cabin crew, comfortable seating, and seamless stopovers at Changi Airport, ensuring a premium travel experience.
3	3_sq_always_flying_crew	sq, always, flying, crew, service, flight, experience, good, best, great	This topic emphasizes Singapore Airlines' consistent delivery of top-notch service, with particular praise for its professional and attentive cabin crew. Passengers value the airline's reliability, modern aircraft, and in-flight comfort, making it a preferred choice for frequent flyers.
4	4_best_airline_singapore_service	best, airline, singapore, service, food, long, entertainment, fly, always, good	This topic underscores Singapore Airlines' reputation as one of the best airlines, with consistent praise for its exceptional service, high-quality food, comfortable seating, and extensive in-flight entertainment. Passengers appreciate its reliability and value for both short and long-haul flights.
5	5_staff_crew_helpful_effient	staff, crew, helpful, friendly, efficient, cabin, check, ground, polite, attentive	This topic focuses on the helpfulness and attentiveness of Singapore Airlines' staff and crew. Passengers commend the team's professionalism, efficiency, and warm demeanor, consistently enhancing the overall travel experience through exceptional service and support.
6	6_seat_good_flight_comfortable	seat, good, flight, comfortable, room, leg, plenty, good, singapore, tall	This topic highlights the comfort and spaciousness of Singapore Airlines' seating across various classes. Passengers frequently mention ample legroom, well-designed seats, and quality in-flight amenities, contributing to a relaxing and enjoyable travel experience, even on long-haul flights.
7	7_movie_film_great_good	movie, film, great, good, choice, entertainment, selection, friendly, food, music	This topic highlights the extensive and high-quality in-flight entertainment options offered by Singapore Airlines. Passengers consistently praise the wide selection of movies and films, alongside excellent service and good food, making long-haul flights enjoyable and engaging.
8	8_food_staff_good_service	food, staff, good, service, comfortable, excellent, friendly, great, crew, meal	Similar to Topic 0, this topic highlights the quality of food and service provided by Singapore Airlines. However, it places additional emphasis on the staff's attentiveness and friendliness, enhancing passengers' comfort and satisfaction.
9	9_economy_premium_seat_extra	economy, premium, seat, extra, worth, leg, room, zealand, money, service	This topic highlights the value of Premium Economy for its extra comfort, including wider seats, more legroom, and enhanced amenities, making it worth the additional cost. Passengers also appreciate the consistent service quality in both Economy and Premium Economy classes.

Table 13: Top 10 topics for negative reviews and key findings

Top 10 Topics for Negative Reviews			
Topic	**Topic label**	**Keywords**	**Key findings**
0	0_seat_bed_leg_flat	seat, bed, flat, leg, uncomfortable, foot, side, airbus, new, wide	This topic highlights passenger dissatisfaction with the design and comfort of seats, particularly in business class. Complaints include cramped footwells, hard seat surfaces, and awkward sleeping positions. While service and food quality remain strong, seat comfort negatively impacts the overall experience.
1	1_food_flight_service_good	food, flight, service, good, singapore, airline, expectation, didnt, disappointed, flown	Passengers express disappointment in food quality, uncomfortable seating, and inconsistent service on Singapore Airlines. Complaints include inedible meals, limited beverage options, poor service responsiveness, and cramped seating. High expectations for a five-star airline are often unmet, leading to dissatisfaction and a perception of declining standards.
2	2_seat_flight_meal_class	seat, flight, meal, class, passenger, one, business, lounge, economy, fly	Similar to Topic 0, this topic addresses dissatisfaction with seat comfort and inadequate amenities across classes. Passengers criticize cramped seating, poor meal quality, and inconsistent service, particularly in premium classes, leading to unmet expectations for a globally reputed premium airline.
3	3_economy_premium_seat_class	economy, premium, seat, class, extra, qantas, toilet, much, better, worth	Similar to Topic 2, this topic highlights dissatisfaction with Singapore Airlines' Premium Economy. Passengers criticize narrow seats, shared amenities with Economy, mediocre meals, and lack of exclusive services. Comparisons with competitors like Qantas and Virgin Atlantic reveal a perceived decline in value and service quality.
4	4_luggage_bag_baggage_lost	luggage, bag, baggage, lost, day, damaged, suitcase, call, damage, missing	This topic highlights frequent complaints about lost, delayed, or damaged luggage on Singapore Airlines. Passengers express frustration over inadequate compensation, poor communication, and inconsistent handling of issues, significantly impacting their travel experience and trust in the airline's service reliability.
5	5_refund_cancelled_bank_ticket	refund, cancelled, bank, ticket, money, covid, month, email, cancellation, refunded	This topic highlights significant frustration with delayed or unprocessed refunds for cancelled flights. Similar to Topic 4, passengers report poor communication, generic responses, and a lack of accountability. These issues erode trust in Singapore Airlines, particularly during high-stress situations involving cancellations.
6	6_baby_seat_bassinet_child	baby, seat, bassinet, child, crew, birthday, cake, special, son, name, kid,	This topic emphasizes mixed experiences when traveling with children on Singapore Airlines. While many praise attentive crew and helpful services for young passengers, others criticize inadequate seating arrangements, lack of children-specific meals, and disruptions caused by poorly managed child seating policies.
7	7_flight_delayed_delay_hour	flight, delayed, delay, hour, pm, airport, connection, next, missed, connecting	This topic highlights dissatisfaction with flight delays and their poor management by Singapore Airlines. Passengers report missed connections, inadequate compensation, lack of clear communication, and unsatisfactory customer support during disruptions, leading to frustration and a diminished travel experience.
8	8_ticket_change_date_booking	ticket, change, date, booking, refund, customer, pay, original, booked, rebook	Passengers report significant frustration with ticket changes and rebookings on Singapore Airlines. Similar to Topic 5, issues include unclear policies, unexpected fees, poor communication, and unhelpful customer service. These experiences often leave travelers feeling misled and dissatisfied with the airline's handling of changes.
9	9_sia_service_always_year	sia, service, always, year, food, best, price, flight, flying, standard	This topic highlights long-term passengers' concerns about a decline in Singapore Airlines' service quality compared to pre-COVID standards. Similar to Topic 7, issues include reduced food quality, poor ground service, and inconsistent customer interactions. Many feel rising prices no longer justify the diminished experience.

Table 12 and 13 present the top 10 topics for positive and negative reviews, illustrating both coherence and differentiation. Many topics cover similar themes, such as "food" and "service," but BERTopic's granularity allows further segmentation based on specific influencing factors. For instance, while Topics 0 and 8 in positive reviews both highlight food quality, Topic 0 emphasizes broader airline service, while Topic 8 focuses on comfort and staff attentiveness. This distinction underscores BERTopic's sensitivity to subtle differences within the same thematic domain, providing actionable insights for enhancing specific service areas.

Although BERTopic provides detailed topic segmentation based on meaning and context, it also presents limitations, with a significant portion of reviews remaining unclassified as Topic -1. This highlights the complexity of customer feedback. To address this, the next section leverages Zero-Shot Classification to refine and assign specific topics to these ambiguous reviews, enhancing actionable insights.

Refinement of Topic -1 Using Zero-Shot Classification

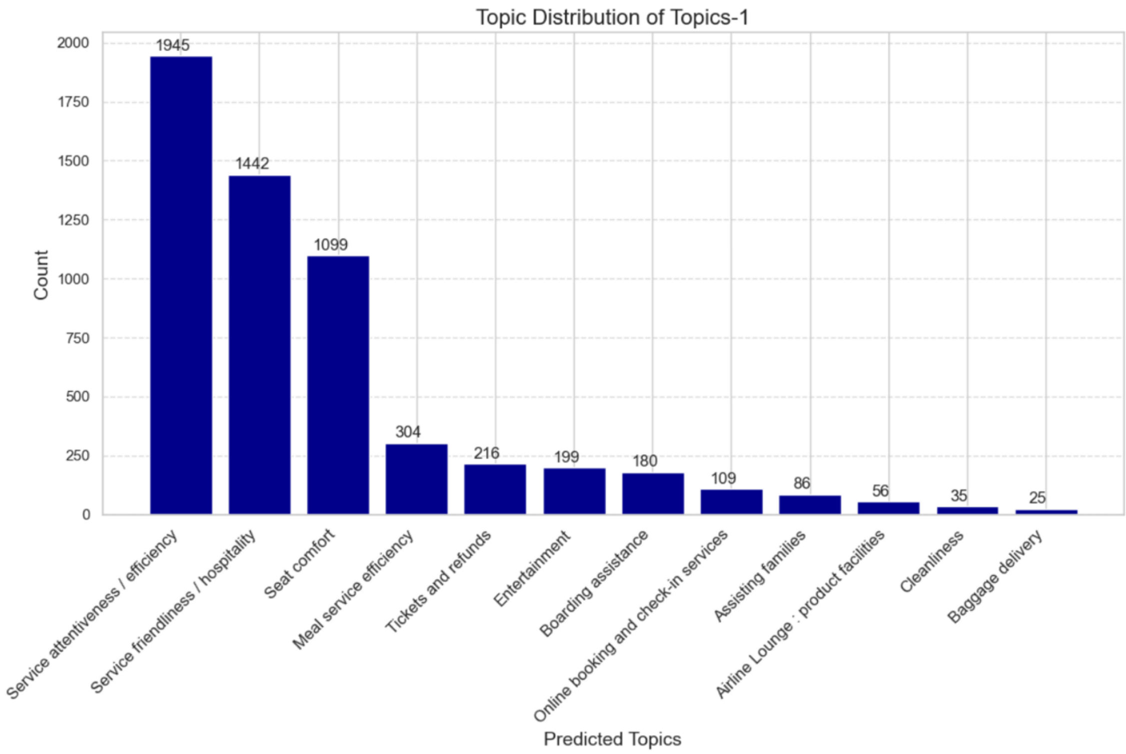


Figure 25: Topic Distribution of Topic-1

Zero-Shot Classification was applied to reviews categorized as Topic -1 (5,688 reviews) by BERTopic, successfully reducing ambiguity and aligning them with predefined themes. Among these, 188 reviews (3.3%) with low classification confidence were grouped under the General Topic, reflecting cases where the content lacked sufficient specificity to assign a definitive theme.

The results (Figure 25) revealed a strong emphasis on service-related attributes within reviews categorized under Topic -1. Among these, Service attentiveness/efficiency (1,945 reviews) and Service friendliness/hospitality (1,442 reviews) were most frequently identified, reflecting the central role of efficient and personalized service in customer perceptions. Additional themes such as Seat comfort (1,099 reviews) and Meal service efficiency (304 reviews) further underscore the importance of physical comfort and operational excellence.

Interestingly, topics such as Cleanliness (56 reviews) and Baggage delivery (25 reviews) were mentioned far less frequently, potentially indicating these areas are either less critical or already meeting customer expectations. However, it is critical to highlight that these findings represent an aggregate view of reviews reclassified from Topic -1 without differentiation by sentiment. As such, the themes identified here may include both positive and negative feedback, necessitating further sentiment segmentation for actionable insights.

This refined distribution demonstrates Zero-Shot Classification's ability to uncover meaningful patterns within unstructured data, bridging the gap left by BERTopic. In the subsequent section, the results of Zero-Shot Classification will be integrated with the broader thematic analysis from BERTopic and sentiment to create a unified perspective, enabling comprehensive insights and actionable recommendations for service improvement.

Exploratory Data Analysis (EDA) of Sentiment Analysis

The result is obtained by mapping of reviews to Skytrax criteria was achieved using a hybrid approach combining BERTopic and Zero-Shot Classification outputs.

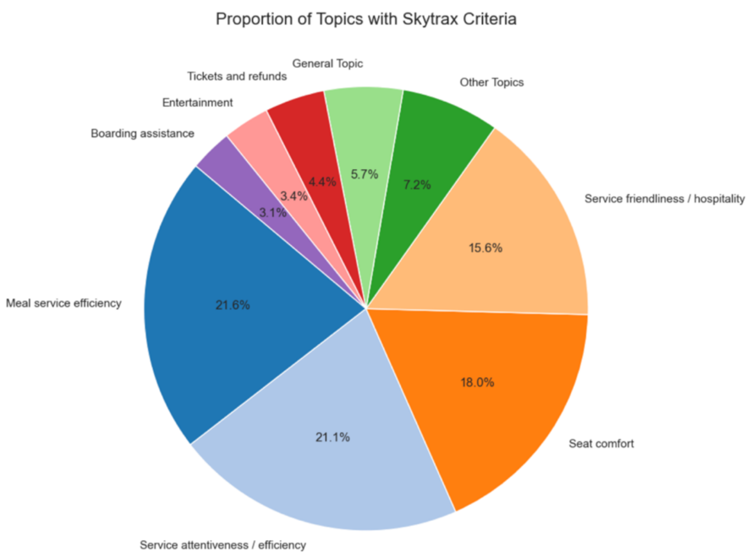


Figure 26: Proportion of Topics

Figure 26 provides an overview of the distribution of topics, highlighting that Meal service efficiency (21.6%), Service attentiveness/efficiency (21.1%), and Seat comfort (18.0%) dominate customer reviews, showcasing the importance of operational and comfort-related attributes.

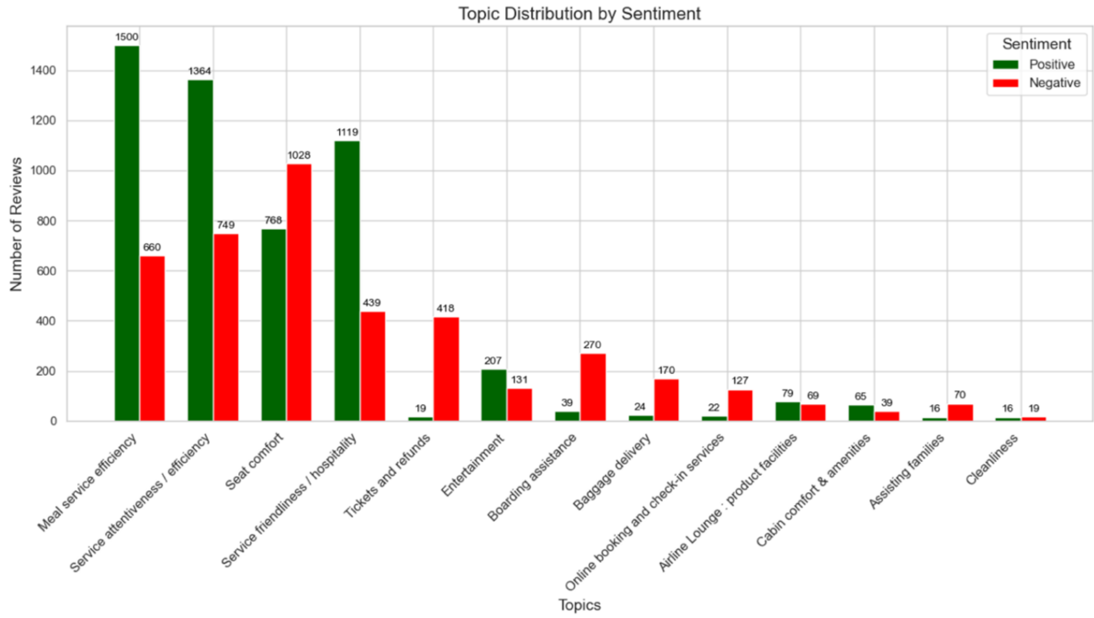


Figure 27: Topic Distribution by Sentiment

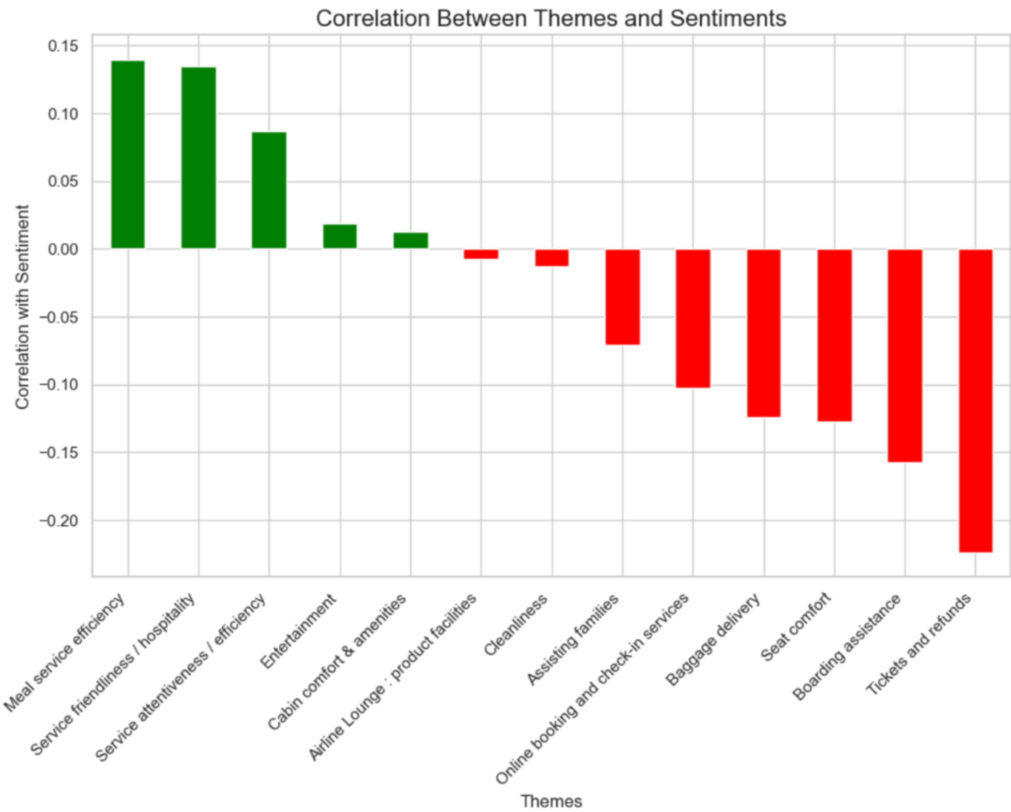


Figure 28: Correlation Between Themes and Sentiments

From Figure 27, Meal service efficiency and Service attentiveness/efficiency stand out with the highest number of positive reviews, at 1,500 and 1,364 reviews, respectively. These findings underscore the airline's operational excellence and attentive service as pivotal strengths in shaping a premium customer experience. Similarly, Service friendliness/hospitality follows with 1,119 positive reviews, emphasizing the importance of personalized interactions in enhancing customer satisfaction. These results affirm Singapore Airlines' positioning as a leader in in-flight service delivery.

Adding depth to this analysis, Figure 28 reveals that these themes—Meal service efficiency (0.1396), Service friendliness/hospitality (0.1349), and Service attentiveness/efficiency (0.0866)—not only attract high volumes of positive feedback but also exhibit the strongest positive correlations with sentiment. This alignment between volume and correlation reinforces their significance in driving customer satisfaction and sustaining brand loyalty.

Conversely, challenges emerge from themes with high negative feedback volumes and strong negative correlations. Tickets and refunds and Baggage delivery, which received 418 and 170 negative reviews, respectively (Figure 27), also show the strongest negative correlations with sentiment at -0.2237 and -0.1244 (Figure 28). These findings highlight procedural inefficiencies and unmet customer expectations in these domains. Similarly, Boarding assistance (-0.1570) and Seat comfort (-0.1272) show high negative sentiment ratios and correlations, indicating a pressing need for targeted interventions to address these pain points.

Interestingly, themes such as Entertainment (0.0188) and Cabin comfort & amenities (0.0129) exhibit neutral or slightly positive correlations, suggesting relatively balanced feedback. While customer expectations in these areas might already be met to a satisfactory level, even minor enhancements could bolster the overall travel experience.

By integrating insights from both figures, a clear dual strategy emerges. Sustaining excellence in high-performing areas, such as service attentiveness and meal quality, is essential to maintaining competitive differentiation. Simultaneously, prioritizing improvements in weaker areas like ticketing processes, baggage handling, and boarding services is critical to reducing dissatisfaction and strengthening customer loyalty. These findings provide a roadmap for strategic investments to reinforce Singapore Airlines' industry leadership.

The subsequent section will expand on these insights through a temporal analysis of customer feedback, examining how sentiment and expectations evolve across different periods, such as pre- and post-pandemic phases. This dynamic perspective will enable the airline to align its service strategies proactively with shifting customer priorities, ensuring sustained excellence in the long term.

Insights Across Temporal Phases

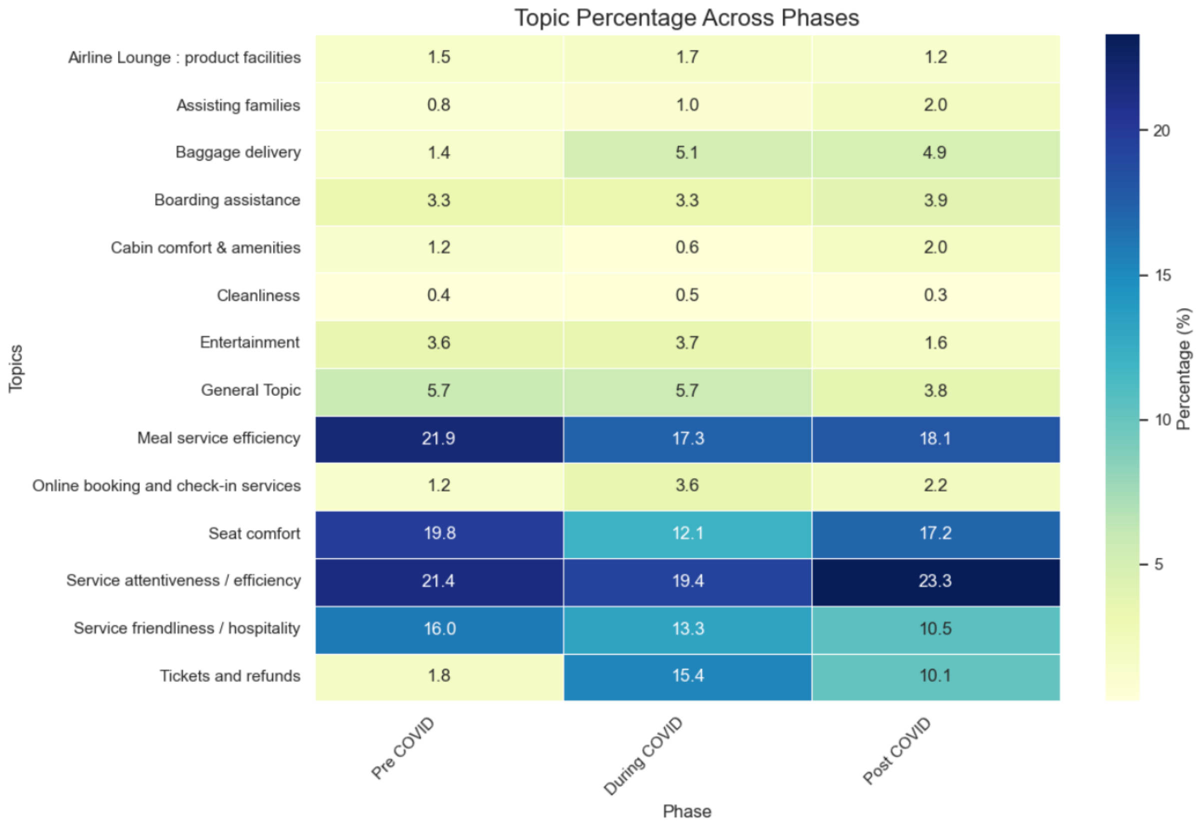


Figure 29: Heatmap of Topic Percentage Across Phases

Figure 29 offers a comprehensive overview of the distribution of customer reviews across different themes and temporal phases. This visualization provides insight into the shifts in customer focus and highlights the dynamic nature of customer expectations in response to external factors such as the pandemic.

A significant observation from the data is the dramatic increase in reviews related to Tickets and refunds during the COVID-19 period. Before the pandemic, this topic accounted for a mere 1.8% of the reviews, reflecting its relatively minor role in customer discourse. However, during the pandemic, its proportion surged to 15.4%, likely driven by widespread travel disruptions and a surge in refund and ticket amendment requests caused by government-imposed travel restrictions and quarantine requirements (International Air Transport Association, 2021). Although this proportion decreased to 10.1% in the Post-COVID phase, it remains notably higher than in the Pre-COVID period, signaling the continued importance of efficient and transparent ticketing processes in maintaining customer trust.

Other topics, such as Seat comfort and Service attentiveness/efficiency, show relatively stable proportions over time. However, the notable decrease in Seat comfort during COVID, from 19.8% to 12.1%, followed by a rebound to 17.2% post-pandemic, suggests the impact of reduced passenger volumes due to travel restrictions. This decrease may reflect lower customer expectations for comfort during this period, when stringent public health measures significantly altered the travel experience (Civil Aviation Authority, 2020). While the heatmap does not explicitly indicate sentiment, it is plausible that the challenges of limited travel options and reduced capacity influenced customer discourse on comfort-related issues.

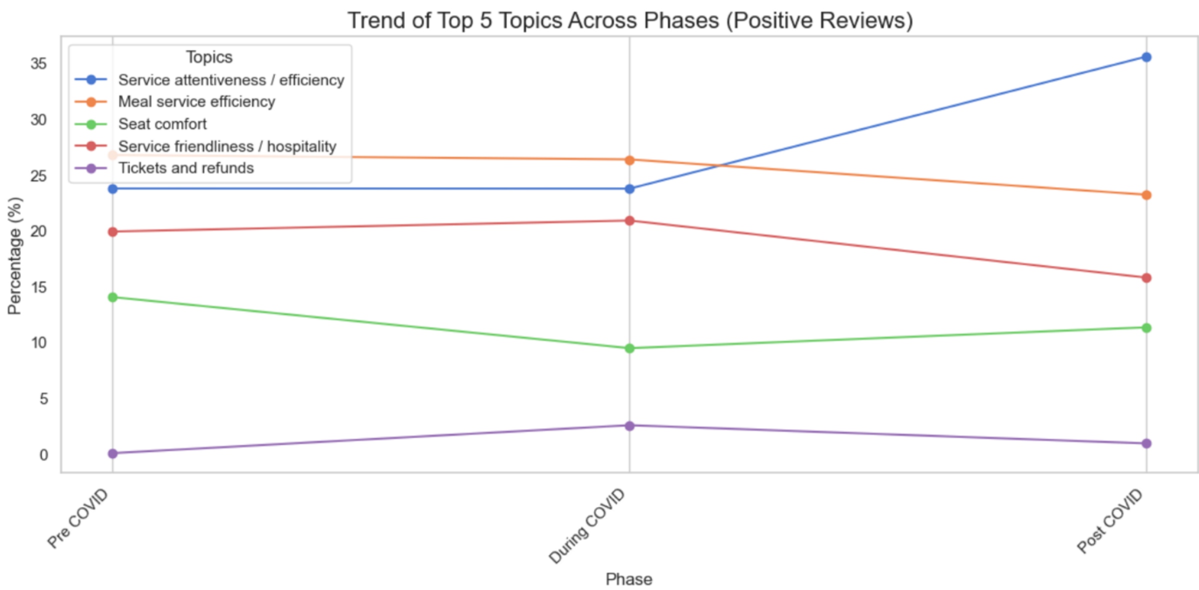


Figure 30: Trends in Positive Sentiment

The analysis of positive sentiment trends reveals Singapore Airlines' key strengths in maintaining exceptional service quality across phases. The consistent prominence of Service attentiveness/efficiency and Meal service efficiency highlights the airline's ability to sustain operational excellence and meet customer expectations, even under the constraints of the pandemic. While Service friendliness/hospitality experienced a slight decline during COVID-19, likely due to reduced in-flight interactions from health protocols, it remains a cornerstone of positive customer feedback. These findings reaffirm the importance of focusing on these core service areas to strengthen customer loyalty and sustain Singapore Airlines' premium positioning.

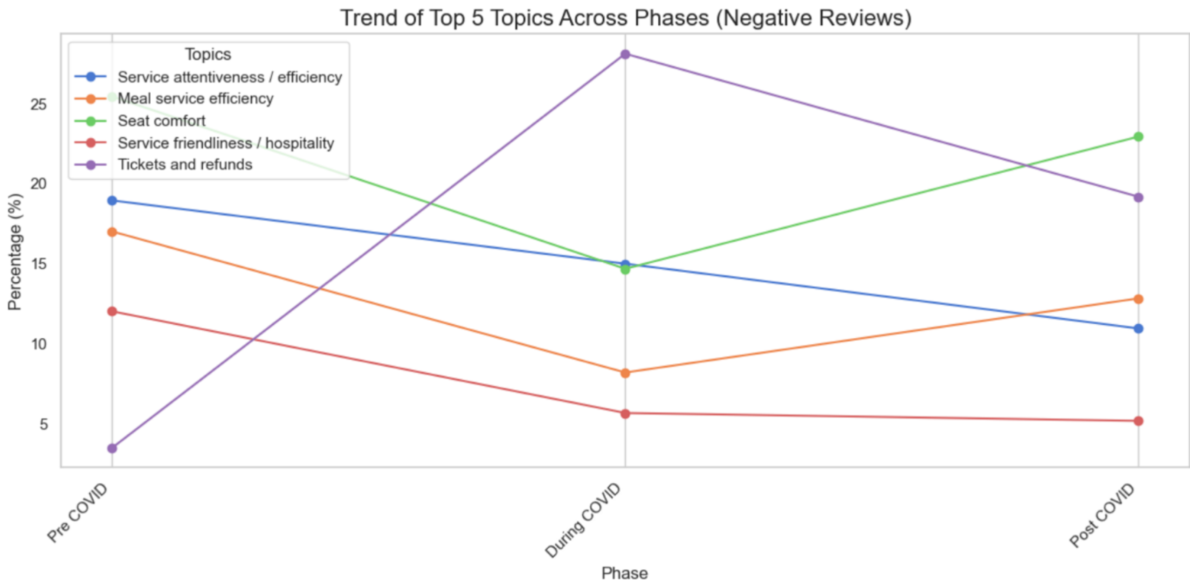


Figure 31: Trends in Negative Sentiment

The analysis of negative sentiment trends underscores areas requiring improvement, with Tickets and refunds emerging as the most significant negative theme during the COVID-19 phase, accounting for 27.8% of all negative reviews. Although its proportion decreased to 16.1% in the Post-COVID phase, it remains a critical concern, reflecting ongoing procedural inefficiencies or unmet expectations in ticketing and refund policies. Addressing this issue should be a priority for Singapore Airlines to enhance customer satisfaction and mitigate dissatisfaction. The data also shows a gradual increase in negative sentiment related to Seat comfort, particularly in the Post-COVID phase, which coincides with the resumption of high passenger volumes. This trend highlights the importance of addressing customer concerns regarding cabin design and personal space.

Strategic Recommendations for Service Enhancement

The findings from this analysis underline the need for Singapore Airlines to adopt a dual strategy, focusing on both sustaining excellence in key areas and addressing critical customer pain points.

First, to address the heightened risk perception among travelers in the post-COVID era, SIA should implement strategies that mitigate financial, psychological, and health-related concerns, which significantly influence purchasing decisions (Seabra et al., 2014; Suau-Sanchez et al., 2020). Flexible booking policies, such as refundable tickets and fee waivers for changes or cancellations, could alleviate financial risks, encouraging customers to book flights with confidence (Bettman, 1973). Offering travel insurance options that cover pandemic-related disruptions would further enhance consumer trust, aligning with Xu et al.'s (2019) findings that immediate and tangible compensations, such as monetary refunds or upgrades, are critical in reducing dissatisfaction, particularly for full-service airlines like SIA.

Second, Investment in digital transformation should also be a priority to address critical pain points that emerged during the pandemic, such as ticketing and refund processes. Seamless and efficient digital platforms can improve operational efficiency, build customer trust, and enable proactive responses to disruptions. Advanced analytics tools, supported by real-time data, can streamline communication channels and decision-making, ensuring transparency and minimizing uncertainties during the booking process (Pereira, 2023; Song et al., 2020).

Third, attention should be given to addressing concerns related to Seat comfort. The gradual increase in negative feedback in the Post-COVID phase suggests that efforts to redesign cabins with greater emphasis on personal space and ergonomic seating could mitigate dissatisfaction. Investing in flexible cabin configurations that cater to diverse customer needs may also enhance satisfaction.

Finally, Singapore Airlines should continue its investment in staff training to maintain high standards in Service attentiveness/efficiency and Meal service quality, which have consistently contributed to positive customer sentiment. Proactive measures to enhance personalized interactions and ensure consistent meal delivery will further reinforce the airline’s premium positioning.

These recommendations, derived from sentiment analysis and thematic mapping, provide a strategic framework for Singapore Airlines to address evolving customer expectations and strengthen its industry leadership. By integrating these insights with broader temporal trends, the airline can dynamically align its service strategy with customer needs and maintain its competitive edge.

Predictive and Forecast Modelling of Customer Sentiments (RQ4)

Model Selection for Forecasting

To predict trends in customer sentiments for the most frequently reviewed topics—Service attentiveness/efficiency, Meal service efficiency, and Seat comfort—two advanced time-series forecasting models, Long Short-Term Memory (LSTM) and Prophet, were selected. These models were trained on historical sentiment data to anticipate future trends within both positive and negative sentiment categories, as illustrated in Figure 32.

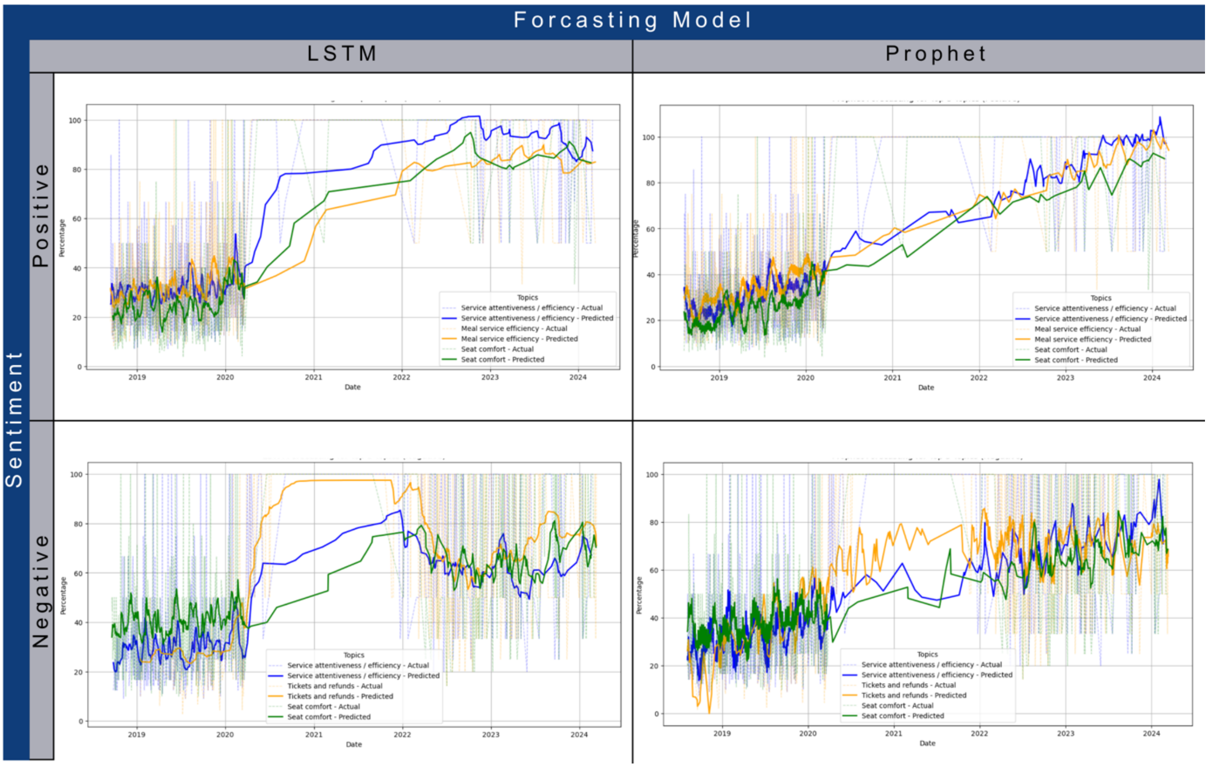


Figure 32: Comparison of Model Training

Model Evaluation and Selection

The accuracy of both models was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). RMSE provides an estimate of the standard deviation of prediction errors, with a focus on penalizing larger errors. MAE offers an average magnitude of prediction errors, making it less sensitive to extreme deviations compared to RMSE. MAPE expresses prediction errors as a percentage of actual values, enabling comparisons across varying data scales.

Table 14: Model Evaluation Scores

Model	Topic	RMSE	MAE	MAPE
Prophet	Service attentiveness / efficiency	17.8022	13.3020	92.2243
	Meal service efficiency	17.5846	13.6155	110.4600
	Seat comfort	16.4005	12.0420	112.3094
LSTM	Service attentiveness / efficiency	17.7774	13.3973	44.9073
	Meal service efficiency	18.2435	13.9485	48.3495
	Seat comfort	17.1119	12.4068	57.1505

As detailed in Table 14, LSTM consistently demonstrated superior performance compared to Prophet. For instance, LSTM achieved a lower MAPE score for Seat comfort (57.1505) compared to Prophet (112.3094), reflecting its ability to capture data variability and predict trends with greater accuracy. Similar patterns were observed across the other topics, with LSTM also outperforming Prophet in RMSE and MAE metrics. The results indicate that LSTM's ability to handle non-linearities in the data provides a more reliable framework for sentiment forecasting. Consequently, LSTM was selected as the preferred model for predicting future sentiment trends.

Analysis of Forecasting Results

The results of the LSTM model highlight critical trends in customer sentiment across positive and negative categories.

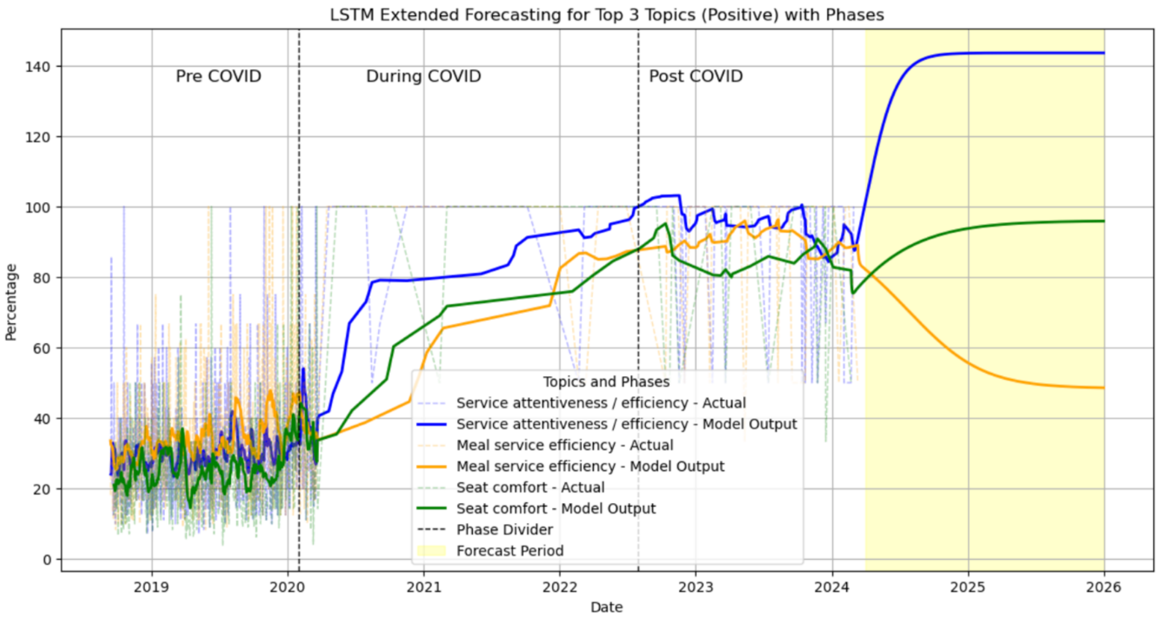


Figure 33: Forcasting Trend of Top 3 Topics in Positive Sentiment

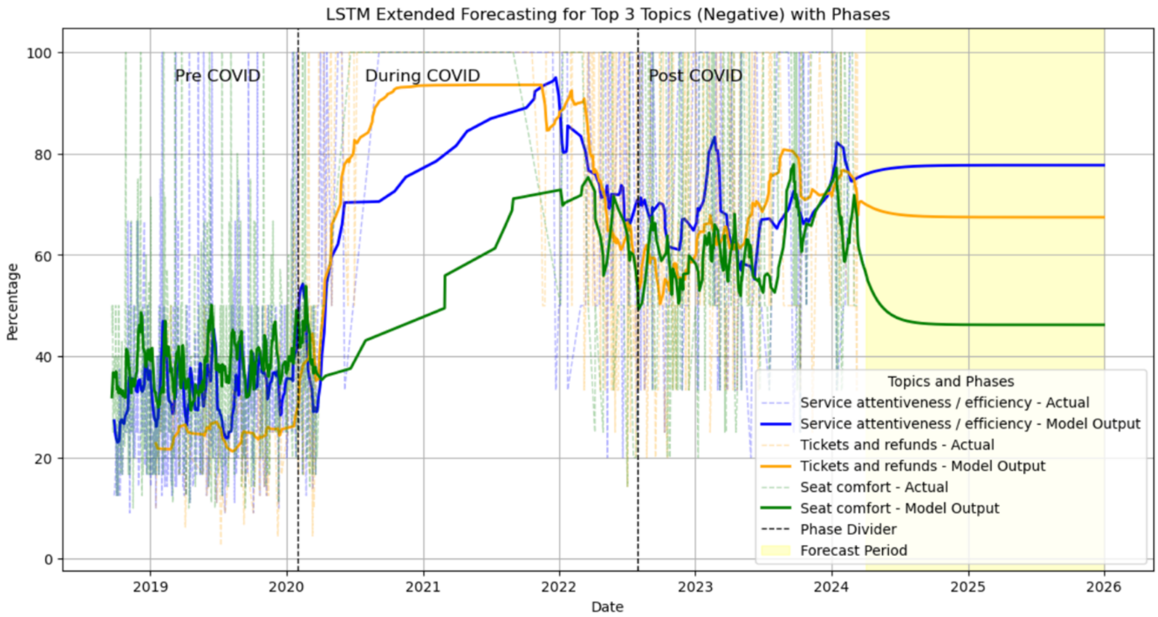


Figure 34: Forcasting Trend of Top 3 Topics in Negative Sentiment

According to Figure 33 and Figure 34, the analysis of Service attentiveness/efficiency and Seat comfort reveals interconnected trends across both positive and negative sentiments, providing valuable insights into customer feedback dynamics.

For Service attentiveness/efficiency, the forecasts for positive sentiment indicate a steady upward trajectory, with projections suggesting continued growth through 2025. This trend underscores Singapore Airlines' sustained focus on delivering personalized and responsive service, which remains a core strength. However, the forecasts for negative sentiment show a marginal increase, emphasizing the need for consistent service training and process refinement to mitigate potential lapses. The simultaneous growth in positive mentions and the slight increase in complaints may indicate that while service remains a key strength, higher customer expectations post-pandemic may also magnify any inconsistencies.

For Seat comfort, the relationship between positive and negative sentiment trends is particularly revealing. Positive sentiment forecasts show stabilization at higher levels compared to the pandemic era, reflecting ongoing efforts to enhance passenger comfort through measures such as improved seating design and cabin configurations. This aligns with the decline in negative feedback for Seat comfort, suggesting that operational enhancements are addressing long-standing concerns. The pandemic's impact on reduced passenger volumes may have temporarily alleviated discomfort, as fewer passengers resulted in more available space. However, as travel volumes normalize, these improvements have prevented a return to pre-pandemic complaint levels, indicating effective adjustments to meet evolving customer expectations.

Meal service efficiency exhibits a nuanced trend, with positive feedback projected to decline slightly by 2025. This could reflect challenges in maintaining consistency in meal quality and variety as the airline scales its operations post-pandemic. Interestingly, the absence of significant negative sentiment for this theme suggests that while improvements are possible, the overall service level remains satisfactory for most passengers. Future efforts could focus on innovation in meal offerings to sustain customer interest and satisfaction.

Tickets and refunds continue to stand out as a critical area for improvement. Although negative sentiment is forecasted to decline gradually, the prominence of this issue during the pandemic underscores the operational challenges faced during periods of disruption. The airline's efforts to streamline refund processes and enhance transparency are likely contributing to the reduction in complaints, but further refinements are necessary to build customer trust in this area.

The pandemic has had a profound influence on customer expectations and feedback patterns. For Service attentiveness/efficiency, heightened health and safety protocols during the pandemic might have led to stricter evaluations by passengers, while for Seat comfort, reduced passenger volumes likely softened perceptions of cramped spaces. Post-pandemic, as volumes recover, these factors highlight the importance of addressing both operational challenges and evolving expectations to maintain customer satisfaction.

In summary, the interconnected trends between positive and negative sentiments underscore the importance of adopting a balanced strategy. Sustaining excellence in areas like Service attentiveness/efficiency and Seat comfort while addressing specific challenges in Tickets and refunds offers a pathway to enhancing customer satisfaction. The insights from the pandemic period provide valuable lessons on the impact of external disruptions on passenger expectations and service perceptions, offering a blueprint for resilience in future challenges.

Limitations of the Study

This study acknowledges several limitations that may impact the robustness of its findings. DistilBERT, while effective in contextual understanding, tends to overrepresent positive sentiments, particularly when datasets are imbalanced. For instance, sarcasm in statements like "just loved the cramped seats" is often misclassified as positive due to limitations in detecting nuanced tones (Dorssers, 2024). To address this, the study employs compound scores (greater than 0 for positive, less than 0 for negative) to improve sentiment classification accuracy. Fine-tuning DistilBERT on balanced datasets and applying cross-validation could further enhance its ability to handle subtle sentiments.

In thematic analysis, Zero-Shot Classification relies heavily on predefined labels, which may fail to capture all emerging or overlapping themes. To mitigate this, the study uses confidence scores to evaluate theme assignments, prioritizing themes with scores above 0.6. Reviews with lower scores are categorized as "General Topics," reflecting high variability. While this ensures greater clarity, overlapping or ambiguous themes may still limit the model's performance. Iterative refinement of labels and active learning could improve adaptability and accuracy (Ramalingam et al., 2023).

BERTopic, used for thematic analysis, struggles with short or fragmented reviews that lack context, which can dilute thematic clarity. For instance, reviews like "great staff, poor food" may result in overemphasis on frequently occurring words rather than thematic balance (Majumder et al., 2019). Combining confidence scores with domain-specific stopword lists helps reduce this bias but does not fully eliminate it.

The forecasting models, LSTM and Facebook Prophet, rely on historical trends and assume continuity, making them less effective for predicting abrupt changes like the COVID-19 pandemic. Incorporating external variables, such as economic conditions or competitor actions, could enhance these models' ability to adapt to dynamic shifts in customer behavior (Product Teacher, n.d.).

Lastly, the dataset is sourced exclusively from TripAdvisor, which may introduce bias due to its tendency to attract extreme opinions, either highly positive or negative. Moderate experiences may be underrepresented (citation missing). Including data from platforms like Skytrax or Flight-Report could provide a more balanced perspective. Additionally, the temporal scope of 2018 to

2024, while insightful, does not capture long-term trends. Expanding the timeframe and integrating broader datasets would enhance the generalizability of future findings.

Despite these limitations, the study's use of compound and confidence scores, combined with advanced modeling techniques, enhances the reliability of its results and highlights areas for improvement.

Ethical Implications

This study adheres to rigorous ethical standards throughout the data lifecycle, from collection to analysis and reporting. The dataset comprises 10,000 anonymized TripAdvisor reviews, systematically collected using Python-based web scraping tools for the period 2018–2024. TripAdvisor's review guidelines ensure reviews are based on first-hand traveler experiences, enhancing data reliability and authenticity (Tripadvisor, 2024). The collection process included removing duplicates, irrelevant content, and any data compromising integrity. All personally identifiable information (PII) was anonymized in strict adherence to GDPR and data confidentiality principles (European Parliament, 2023).

To address potential biases in user-generated content, such as the overrepresentation of extreme opinions or imbalanced data, this study employed multiple result reviews to identify and mitigate patterns of bias. Through iterative evaluations of outputs from models such as DistilBERT and BERTopic, potential issues were identified, allowing refinements to improve structural accuracy and reduce the influence of extreme sentiments. Additionally, data sampling was conducted to ensure a balanced distribution of positive and negative sentiments, thereby enhancing the reliability of the analysis. These measures aimed to minimize the inherent biases of user-generated reviews while maintaining the integrity of insights. Future work could further strengthen these efforts by implementing advanced sensitivity analyses, such as testing model robustness under varying parameters or introducing balanced reweighting strategies to refine sentiment classification and thematic accuracy.

The anonymized dataset is securely stored on the University of Exeter OneDrive, encrypted and accessible only to authorized personnel. Data will be retained for 4 months for verification purposes and securely destroyed afterward, adhering to institutional policies.

By integrating TripAdvisor's guidelines, GDPR, and ethical AI principles, this study ensures privacy, minimizes bias, and upholds academic and ethical standards.

Conclusion

Summary of Key Findings

This study analyzed the evolution of customer sentiment toward Singapore Airlines across three distinct periods—pre-COVID, during COVID-19, and post-COVID—using 10,000 anonymized TripAdvisor reviews. By leveraging advanced methodologies such as DistilBERT for sentiment analysis, BERTopic for thematic exploration, and predictive models like LSTM and Facebook Prophet, the research offered valuable insights into factors shaping customer satisfaction. Aligning the findings with industry benchmarks like the Skytrax World Airline Awards provided a comprehensive understanding of how disruptions and service adaptations influenced customer perceptions over time.

Pre-COVID reviews reflected overwhelmingly positive sentiments, driven by Singapore Airlines' high service standards, exceptional in-flight service, and attentive customer support, solidifying its reputation as a global leader in customer satisfaction. However, during the COVID-19 period, negative feedback increased due to travel disruptions, policy uncertainties, and communication delays. Passengers expressed frustrations with inconsistent updates and delayed responses but commended proactive safety measures such as enhanced cleanliness protocols and onboard health precautions. In the post-COVID period, sentiment trends gradually improved as services returned to pre-pandemic levels. Passengers praised the reinstatement of full in-flight offerings, including high-quality meals and premium cabin experiences, signaling a recovery in customer satisfaction.

Thematic analysis further illuminated the key drivers of customer sentiment. In-flight service consistently emerged as a crucial factor, with passengers lauding the professionalism and hospitality of cabin crew. The quality of seating and amenities, particularly in premium travel classes, was also a prominent positive aspect. In contrast, dissatisfaction often stemmed from perceived discomfort in economy seating and service inconsistencies. These findings underscored the importance of maintaining uniform service quality across all classes.

Communication and customer support, particularly during periods of disruption, were critical themes throughout the study. Passengers valued timely and empathetic responses, especially when facing cancellations or policy changes. Negative feedback frequently pointed to delays in refund processing and inconsistent communication during the pandemic, emphasizing the need for better customer relationship management and faster resolution of issues.

Forecasting analyses provided insights into anticipated customer trends. The findings emphasized that maintaining service attractiveness and efficiency, particularly in aspects like meal quality and variety, will be critical to sustaining customer satisfaction. Passengers consistently value high standards in in-flight meals, and enhancing both the quality and diversity of food options across all travel classes could significantly strengthen the airline's appeal. Additionally, operational excellence in handling disruptions, such as efficient ticket adjustments and timely refunds, will remain pivotal for retaining customer trust. The research also suggested that improving overall service efficiency and consistency would further reinforce Singapore Airlines' reputation as a reliable and customer-focused airline.

The research also demonstrated how external events like the COVID-19 pandemic reshaped customer priorities, reinforcing the importance of adaptability and resilience. While traditional drivers of satisfaction, such as in-flight service and amenities, regained prominence post-pandemic, the lasting impact of disruptions highlighted the need for robust systems to handle future challenges effectively.

In conclusion, this study emphasizes the critical role of consistent service, proactive communication, and enhanced attention to service efficiency and meal offerings in improving customer satisfaction. By addressing these key areas, Singapore Airlines can maintain its competitive edge and align its offerings with evolving passenger needs, ensuring sustained leadership in the dynamic aviation industry.

Contributions to the Field

This research provides valuable contributions to the fields of aviation and service management by analyzing customer satisfaction using advanced methodologies such as sentiment analysis, thematic exploration, and predictive modeling. It highlights the importance of key factors such as in-flight service quality, effective communication, and consistent delivery, offering actionable insights to improve customer experiences.

The study demonstrates how disruptions, including the COVID-19 pandemic, influence customer expectations and underscores the need for airlines to adapt through resilient and responsive strategies. By bridging unstructured customer feedback with structured industry benchmarks like the Skytrax World Airline Awards, the research offers a comprehensive perspective on service quality, enabling data-driven decision-making.

Furthermore, the integration of machine learning and natural language processing provides a robust framework for capturing nuanced customer feedback. This approach helps airlines identify opportunities for innovation, such as improving operational efficiency and enhancing service recovery processes during disruptions.

By enriching the understanding of customer sentiment and its evolution, this study equips stakeholders with the tools to align services with shifting expectations. It advances the field by demonstrating how real-time feedback and predictive insights can inform strategic planning, ensuring long-term competitiveness in the aviation industry.

Recommendations for Future Research

Future research should address several limitations identified in this study to enhance the robustness and generalizability of its findings. Expanding the dataset to include reviews from multiple platforms, such as AirlineQuality or AirlineRatings, would provide a more comprehensive representation of customer experiences and reduce potential biases associated with relying solely on TripAdvisor. Such an expansion could improve the diversity of insights and capture moderate opinions often underrepresented on single platforms.

Additionally, the temporal scope of this study, covering 2018–2024, provides valuable insights into sentiment trends influenced by the COVID-19 pandemic but does not account for long-term shifts or emerging trends beyond this period. Extending the timeframe of analysis could capture evolving customer priorities and improve the forecasting of future expectations. Future studies should also explore alternative forecasting models to address the limitations of LSTM and Facebook Prophet, which may struggle with unprecedented disruptions or rapid behavioral changes.

Further, improving the capacity of sentiment analysis tools like DistilBERT to handle subtle nuances, such as sarcasm and mixed sentiments, would enhance the accuracy of sentiment classification. Advanced models or hybrid approaches could be explored to address these challenges. Finally, incorporating broader demographic and cultural contexts into thematic analysis could yield deeper insights into customer behavior across regions, enabling more tailored service improvements. These steps will build on the current findings and provide actionable recommendations to further align services with dynamic customer needs.

References

Abubakar, A. M., & Ilkan, M. (2016). Impact of online WOM on destination trust and intention to travel: A medical tourism perspective. Journal of Destination Marketing & Management, 5(3), 192–201. https://doi.org/10.1016/j.jdmm.2015.12.005

Afaf. (2024). Sentiment analysis: VADER or TextBlob? Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/01/sentiment-analysis-vader-or-textblob/

Agustia, D., Muhammad, N. P. A., &Permatasari, Y. (2020). Earnings Management, Business Strategy, and Bankruptcy Risk: Evidence from Indonesia. Heliyon. https://doi.org/10.1016/j.heliyon.2020.e03317"

Aryan. (2024). Understanding human feelings with NLP and VADER sentiment analysis. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/06/vader-for-sentiment-analysis/

Bigné, E., Ruiz, C., & Currás-Pérez, R. (2019). Destination appeal through digitalized comments. Journal of Business Research, 101, 447–453. https://doi.org/10.1016/j.jbusres.2019.01.020

Bird, S., Klein, E., & Loper, E. (2009). Natural Language Toolkit: A tutorial introduction. Language Processing, 21(5), 123–140.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, 993–1022.

Chauhan, A. (2023). Dealing with erratic data in time series forecasting: Strategies and algorithms. LinkedIn. https://www.linkedin.com/pulse/dealing-erratic-data-time-series-forecasting-ayush-chauhan

Dada, O. A., Olaleye, S. A., Sanusi, I. T., & Obaido, G. (2021). COVID-19 and airline refunds: An analysis of flight passengers' reviews in North America. 19th International Conference e-Society 2021. https://www.esociety-conf.org/wp-content/uploads/2021/03/03_202101L025.pdf

Damle, A. (2024). Time series forecasting with Facebook Prophet. Medium. https://medium.com/@1998ameya/time-series-forecasting-with-facebook-prophet-4d09d5b05c69

Das, D., Sarkar, A., & Debroy, A. (2022). Impact of COVID-19 on changing consumer behaviour: Lessons from an emerging economy. International journal of consumer studies, 46(3), 692–715. https://doi.org/10.1111/ijcs.12786

Deasie, Inc. (2023). Unstructured data problems: Overcoming common challenges. https://www.deasylabs.com/post/unstructured-data-problems-overcoming-common-challenges

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171–4186. https://doi.org/10.18653/v1/N19-1423

Dorssers, O. P. W. (2024). Sarcasm detection in structured text using DistilBERT: Evaluating the impact of text normalization on model performance [Master's thesis, Tilburg University]. Tilburg University Repository. https://arno.uvt.nl/show.cgi?fid=171929

EUROCONTROL. (2020). Daily traffic variation for states. EUROCONTROL. https://www.eurocontrol.int/

EUROCONTROL. (n.d.). Our data. EUROCONTROL. https://www.eurocontrol.int/our-data

European Parliament. (2023, June 8). EU AI Act: first regulation on artificial intelligence. European Parliament. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first- regulation-on-artificial-intelligence.

FareTrack. (2024). The importance of sentiment analysis in airline revenue management strategies. FareTrack. https://www.faretrack.ai/the-importance-of-sentiment-analysis-in-airline

Fatima, Z., & Rahimi, A. (2024). A review of time-series forecasting algorithms for industrial manufacturing systems. Machines, 12(6), 380. https://doi.org/10.3390/machines12060380

Filieri, R. (2015). What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. Journal of Business Research, 68(6), 1261-1270. https://doi.org/10.1016/j.jbusres.2014.11.006

Frenzel, J. (2024). KB sentiment analysis: Rule-based vs. ML approaches. Medium. https://prof-frenzel.medium.com/kb-sentiment-analysis-rule-based-vs-ml-approaches-984a29503716

Garcia, D., & Wang, W. (2020). Challenges in analyzing unstructured data in customer reviews. Journal of Data Science, 18(4), 567-582.

GlobalData. (n.d.). Top 10 travel intermediaries in Asia-Pacific in 2021 by revenue. GlobalData. https://www.globaldata.com/companies/top-companies-by-sector/travel-and-tourism/asia-pacific-travel-intermediaries-by-revenue/

Google. (2024). Get Google reviews. https://support.google.com/business/answer/3474122?hl=en

Google. (2024). Understanding your flight and booking options. Google Travel Help. https://support.google.com/travel/answer/11583641?hl=en

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. ArXiv. https://arxiv.org/abs/2203.05794

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hugging Face. (2024). DistilBERT model documentation. Huggingface. https://huggingface.co/docs/transformers/en/model_doc/distilbert

Hugging Face. (2024). XLM-RoBERTa. Hugging Face. https://huggingface.co/docs/transformers/en/model_doc/xlm-roberta

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International Conference on Web and Social Media (ICWSM) (pp. 216–225). https://ojs.aaai.org/index.php/ICWSM/article/view/14550

International Airport Review. (2024). Skytrax World Airline Awards: Benchmarking excellence in aviation. https://www.internationalairportreview.com

Kanchana. (2024). Singapore Airlines reviews dataset. https://www.kaggle.com/datasets/kanchana1990/singapore-airlines-reviews

Kaur, G., Kaur, A., Khurana, M., & Damaševičius, R. (2024). Sentiment polarity analysis of love letters: Evaluation of TextBlob, VADER, Flair, and Hugging Face transformer. Computer Science and Information Systems, 21(1), 40–40. https://doi.org/10.2298/CSIS240328040K

Komischke, T. (2024). Understanding customer sentiment analysis to drive business growth. CMSWire. https://www.cmswire.com/customer-experience/understanding-customer-sentiment-analysis-to-drive-business-growth/

Lauchlan, S. (2016). Yelp retreats to the US and Canada as international expansion fails. Diginomica. https://diginomica.com/yelp-retreats-to-the-us-homeland-as-international-expansion-fails

Lohmann, G., & Koo, T. T. R. (2013). The impact of airline service quality on passengers' satisfaction and choice behavior. Journal of Air Transport Management, 34, 28–31.

Lou, C., & Yuan, S. (2019). Influencer marketing: How message value and credibility affect consumer trust of branded content on social media. Journal of Interactive Advertising, 19(1), 58–73. https://doi.org/10.1080/15252019.2018.1533501

Majumder, N., Poria, S., Peng, H., Chhaya, N., Cambria, E., & Gelbukh, A. (2019). Sentiment and sarcasm classification with multitask learning. IEEE Intelligent Systems, 34(3), 38–43. https://doi.org/10.1109/MIS.2019.2904691

Moayeri, M., Rabbat, M., Ibrahim, M., & Bouchacourt, D. (2024). Embracing diversity: Interpretable zero-shot classification beyond one vector per class. arXiv. https://arxiv.org/html/2404.16717v1

Naveen, P., & Trojovský, P. (2024). Overview and challenges of machine translation for contextually appropriate translations. iScience, 27(10), 110878. https://doi.org/10.1016/j.isci.2024.110878

Pereira, F., Costa, J. M., Ramos, R., & Raimundo, A. (2023). The impact of the COVID-19 pandemic on airlines' passenger satisfaction. Journal of Air Transport Management, 112, 102441. https://doi.org/10.1016/j.jairtraman.2023.102441

Pontes, D. (2024). What is BERT (Bidirectional Encoder Representations from Transformers)? Zilliz. https://zilliz.com/learn/what-is-bert

Product Teacher. (n.d.). Bidirectional encoder representations from transformers (BERT). Product Teacher. https://www.productteacher.com/quick-product-tips/understanding-bert-for-nlp

Rafalski, K. (2024). 6 must-know Python sentiment analysis libraries. Netguru. https://www.netguru.com/blog/python-sentiment-analysis-libraries

Ramalingam, S., & Krishnan, N. (2023). Enhancing idiom classification on text-based conversational data: Evaluating transformer model performance through k-fold cross-validation. Proceedings of the International Conference on Advanced Electrical Engineering and Computing Innovations (ICAEECI), 1–5. https://doi.org/10.1109/ICAEECI58247.2023.10370954

Rao, V. (2024). LSTM for enterprise time series forecasting. LinkedIn. https://www.linkedin.com/pulse/lstm-enterprise-time-series-forecasting-vasu-rao-sdu2c

Samah, K. A. F. A., Misdan, N. F. A., Jono, M. N. H. H., & Riza, L. S. (2022). The best Malaysian airline companies' visualization through bilingual Twitter sentiment analysis: A machine learning classification. JOIV: International Journal on Informatics Visualization, 6(1), 130–137. https://doi.org/10.30630/joiv.6.1.879

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper, and lighter. arXiv preprint arXiv:1910.01108. https://doi.org/10.48550/arXiv.1910.01108

Singapore Airlines. (2020). Operating statistics – May 2020. https://www.singaporeair.com/content/dam/sia/web-assets/pdfs/about-us/information-for-investors/operating-statistics/opstats-may20.pdf

Singapore Airlines. (2024). Cancellations and refunds. Singapore Airlines. https://www.singaporeair.com/en_UK/bd/travel-info/charges-changes/cancellations-refunds/

Skytrax. (2024). Awards methodology. Skytrax World Airline Awards. https://www.worldairlineawards.com/awards-methodology/

Sobieralski, J. B. (2020). COVID-19 and airline employment: Insights from historical uncertainty shocks to the industry. Transportation Research Interdisciplinary Perspectives, 5, 100123. https://doi.org/10.1016/j.trip.2020.100123

Statista Research Department. (2023). Weekly flights change of global airlines due to COVID-19 as of January 2021. Statista. https://www.statista.com/statistics/1104036/novel-coronavirus-weekly-flights-change-airlines-region/

Statista Research Department. (2024). Weekly aviation jet fuel price 2016–2022. Statista. https://www.statista.com/statistics/526660/aviation-fuel-cost-comparison-across-regions/

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. The American Statistician, 72(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

Tripadvisor. (2024). Review of Singapore Airlines. Tripadvisor. https://www.tripadvisor.com/ShowUserReviews-g1-d8729151-r942167897-Singapore_Airlines-World.html

Tripadvisor. (2024). Review posting guidelines. https://www.tripadvisor.com/Trust-lvBd3L1aU38Y-Review_posting_guidelines.html

Vora, A. (2023). The 10 best customer review sites for businesses in 2024. Semrush Blog. https://www.semrush.com/blog/review-platforms-seo/

Yin, W., Hay, J., & Roth, D. (2019). Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 3914–3923. https://doi.org/10.18653/v1/D19-1404

Youvan, D. (2024). Understanding sentiment analysis with VADER: A comprehensive overview and application. ResearchGate. https://doi.org/10.13140/RG.2.2.33567.98726

Appendices

Appendix A: Singapore Airline Operating Results in May 2020

MAY 2020 OPERATING RESULTS

The operating results for May 2020 are given in the table below.

How SIA Group performed in May 2020	2020	2019	Change
SINGAPORE AIRLINES ("SIA") (PASSENGER)			
Capacity (M seat-km)	481.0	10,937.3	-95.6 %
Passenger-km (M)	44.4	8,729.1	-99.5 %
Passengers carried ('000)	8.6	1,807.3	-99.5 %
Passenger load factor (%)	9.2	79.8	-70.6 pts
Load Factor by Route Region (%)			
East Asia	10.4	76.5	-66.1 pts
Americas	11.7	82.5	-70.8 pts
Europe	7.3	79.3	-72.0 pts
South West Pacific	10.2	81.8	-71.6 pts
West Asia and Africa	-	82.6	n.m.
SILKAIR (PASSENGER)			
Capacity (M seat-km)	4.1	997.5	-99.6 %
Passenger-km (M)	1.3	748.9	-99.8 %
Passengers carried ('000)	0.4	374.1	-99.9 %
Passenger load factor (%)	31.7	75.1	-43.4 pts
Load Factor by Route Region (%)			
East Asia and Pacific	31.7	72.1	-40.4 pts
West Asia	-	82.6	n.m.
FULL SERVICE CARRIERS (SIA & SILKAIR)			
Capacity (M seat-km)	485.1	11,934.8	-95.9 %
Passenger-km (M)	45.7	9,478.0	-99.5 %
Passengers carried ('000)	9.0	2,181.4	-99.6 %
Passenger load factor (%)	9.4	79.4	-70.0 pts
LOW COST CARRIER - SCOOT (PASSENGER)			
Capacity (M seat-km)	67.7	2,762.2	-97.5 %
Passenger-km (M)	1.8	2,357.3	-99.9 %
Passengers carried ('000)	0.6	865.0	-99.9 %
Passenger load factor (%)	2.7	85.3	-82.6 pts
Load Factor by Route Region (%)			
East Asia	3.7	85.1	-81.4 pts
West Asia	-	89.1	n.m.
Rest of World	2.0	83.4	-81.4 pts
GROUP AIRLINES (PASSENGER)			
Capacity (M seat-km)	552.8	14,697.0	-96.2 %
Passenger-km (M)	47.5	11,835.3	-99.6 %
Passengers carried ('000)	9.6	3,046.4	-99.7 %
Passenger load factor (%)	8.6	80.5	-71.9 pts
SIA CARGO			
Capacity (M tonne-km)	360.8	948.0	-61.9 %
Freight tonne-km (M)	270.4	573.1	-52.8 %
Freight carried (M kg)	45.2	107.8	-58.1 %
Cargo load factor* (%)	74.9	60.5	14.4 pts
Load Factor by Route Region (%)			
East Asia	76.9	52.2	24.7 pts
Americas	76.0	61.5	14.5 pts
Europe	76.9	71.3	5.6 pts
South West Pacific	67.5	55.9	11.6 pts
West Asia and Africa	90.2	70.3	19.9 pts

MAY 2020 OPERATING RESULTS

In May 2020, the SIA Group's airlines recorded a 99.6% year-on-year decline in passenger carriage (measured in revenue passenger-kilometres) as travel demand was severely impacted by the Covid-19 pandemic, with border controls and travel restrictions remaining in place around the world. Overall passenger capacity (measured in available seat kilometres) was cut by 96.2% in response. Passenger load factor ("PLF") fell to 8.6%.

SIA's capacity was 95.6% lower compared to last year's, with only a skeletal network in operation connecting Singapore to 14 metro cities. Passenger carriage declined 99.5%, resulting in a PLF of 9.2%. The number of destinations, as well as the frequencies on some existing services, will be increased in June and July 2020, arising from the resumption of transfers via Changi.

SilkAir's passenger carriage decreased by 99.8% against a 99.6% cut in capacity. PLF declined to 31.7%. During the month, SilkAir only operated flights between Singapore and Chongqing.

Scoot's passenger carriage declined 99.9% against a contraction in capacity of 97.5%, which led to a PLF of 2.7%. During the month, Scoot temporarily ceased operations to Southeast Asia, West Asia and Europe, while maintaining flights to Hong Kong and Perth.

Cargo load factor ("CLF") was 14.4 percentage points higher as the capacity contraction of 61.9% outpaced the 52.8% decline in cargo traffic (measured in freight tonne-kilometres). Capacity contraction would have been much greater, save for the deployment of passenger aircraft on cargo-only flights. All regions registered improvements in CLF.

Appendix B: Project Proposal

Comparative Sentiment Analysis of Singapore Airlines' Customer Reviews During Pre-COVID, COVID-19, and Post-COVID Periods Based on TripAdvisor Reviews

Background to the Problem

Singapore Airlines has consistently ranked among the top three airlines in the Skytrax World Airline Awards for over two decades. These awards are widely regarded as a benchmark of excellence in the aviation industry, often referred to as the "Oscars of the aviation industry" due to their rigorous global survey methodology and comprehensive evaluation criteria (International Airport Review, 2024). The Skytrax assessment covers a broad range of categories, including service quality, comfort, in-flight experience, and overall customer satisfaction. High rankings not only enhance an airline’s reputation but also foster consumer trust, as passengers frequently rely on these awards to inform their travel choices. Research indicates that consumers perceive highly ranked airlines as more reliable and superior in quality, significantly influencing their decision-making process (Lohmann & Koo, 2013). Furthermore, airlines use these benchmarks to refine their services, aligning closely with evolving customer expectations.

In addition to formal awards, user-generated reviews on platforms like TripAdvisor provide critical, real-time feedback on customer experiences. TripAdvisor allows passengers to share unfiltered opinions, capturing spontaneous reactions to service encounters. Unlike structured surveys, these reviews offer a more authentic reflection of passenger sentiments, making them a valuable resource for understanding customer satisfaction (Filieri, 2015). The transparency provided by such platforms not only shapes consumer perceptions but also influences their travel decisions. Airlines can leverage these insights to identify areas for improvement and adjust their strategies accordingly.

This study utilizes 10,000 anonymized TripAdvisor reviews collected between August 5, 2018, and March 12, 2024, to analyze how customer perceptions of Singapore Airlines have evolved across three critical periods: pre-COVID, during COVID-19, and post-COVID. By applying sentiment analysis and thematic analysis, the study aims to identify key themes that reflect shifting customer sentiments during these periods. Understanding these trends is essential, as the COVID-19 pandemic has significantly impacted the airline industry, altering customer expectations and service priorities.

By comparing the sentiment-driven insights derived from TripAdvisor with the Skytrax World Airline Awards' assessment criteria, this study aims to provide a comprehensive understanding of customer satisfaction. This dual approach will offer actionable recommendations to help Singapore Airlines enhance its service quality, adapt to changing customer expectations, and maintain its competitive edge in the post-pandemic landscape.

Significance of the Problem

Understanding the key drivers of customer satisfaction is crucial for Singapore Airlines to remain competitive, especially in the aftermath of the COVID-19 pandemic, which has dramatically reshaped customer expectations in the airline industry. By analyzing how customer sentiments have evolved before, during, and after the pandemic, this study aims to uncover critical areas for enhancing service quality and adapting to shifting customer priorities.

The insights derived from this research will not only help Singapore Airlines refine its service offerings but also serve as a benchmark for the broader airline industry. By leveraging advanced techniques such as sentiment analysis and thematic analysis, the study will provide actionable insights into customer perceptions, enabling more effective decision-making. Furthermore, these insights can inform strategies to enhance customer retention, optimize service delivery, and ultimately improve operational efficiency in an increasingly competitive market.

This study's findings would contribute valuable knowledge on how external disruptions like the COVID-19 pandemic influence customer satisfaction, providing airlines with a strategic framework to better anticipate and respond to future challenges.

Aims and Objectives

Aim

The primary aim of this research is to analyze how customer sentiment towards Singapore Airlines has evolved across different periods (pre-COVID, during COVID-19, and post-COVID) and provide data-driven insights to enhance service quality. By examining customer reviews from 2018 to 2024, the study aims to uncover key drivers of satisfaction and areas for improvement, especially in response to significant disruptions like the COVID-19 pandemic.

Objectives

1. **Analyze sentiment trends over distinct time periods (2018-2024):**
   Investigate how customer sentiment has shifted before, during, and after the COVID-19 pandemic, focusing on changes in feedback patterns and customer priorities.
2. **Identify key drivers of customer satisfaction and dissatisfaction across different periods:**
   Use sentiment analysis to determine which service attributes (e.g., meal service efficiency, seat comfort, service friendliness / hospitality) contributed to positive and negative reviews during each phase.
3. **Compare and analyze themes across time periods using thematic analysis:**
   Conduct thematic analysis to identify common topics and emerging concerns in customer reviews across the three time frames, highlighting the impact of COVID-19 on customer expectations.
4. **Forecast future customer sentiment trends:**
   Utilize time-series analysis to predict how customer satisfaction levels in main topics may change in the future based on historical sentiment trends, allowing for proactive service improvements.
5. **Develop actionable recommendations for service enhancement:**
   Based on the analysis of sentiment trends and thematic insights, provide targeted recommendations to improve customer experience and maintain Singapore Airlines' competitive edge.

Problem Statement

Despite consistently ranking among the top three airlines in the Skytrax World Airline Awards, Singapore Airlines faces a continuous challenge in identifying and addressing the specific factors that sustain long-term customer satisfaction. The dynamic nature of customer expectations, particularly in response to significant disruptions such as the COVID-19 pandemic, underscores the importance of leveraging data-driven insights to maintain its competitive edge.

Traditional studies often rely on general customer satisfaction metrics, which may overlook the detailed nuances embedded in user-generated content. This study addresses this gap by analyzing 10,000 anonymized TripAdvisor reviews from 2018 to 2024. By focusing on sentiment analysis segmented into positive and negative categories, the study emphasizes the identification of key areas for service improvement. This binary sentiment classification is particularly valuable in highlighting negative feedback that may otherwise be obscured within mixed reviews, providing actionable insights for enhancing service quality.

To supplement the sentiment analysis, thematic analysis explores recurring topics and evolving priorities across three critical phases: pre-COVID, during COVID-19, and post-COVID. This layered analysis aims to understand how customer experiences and expectations have shifted over time and to map these findings against the structured criteria of the Skytrax World Airline Awards. This approach offers a comprehensive perspective on customer satisfaction, linking real-time user feedback to industry benchmarks.

The results of this study will inform strategic service enhancements by identifying patterns in customer sentiments and themes while forecasting emerging trends. This will enable Singapore Airlines to proactively adapt its offerings to align with evolving customer needs, ensuring its continued success as one of the world's leading airlines.

Research Questions

1. How has customer sentiment evolved across the periods of pre-COVID, during COVID-19, and post-COVID, based on the analysis of customer reviews, and what are the key insights from the binary sentiment classification into positive and negative categories?
2. What themes and topics are strongly associated with positive and negative sentiments, and how do these themes vary across different time periods?
3. How can the themes identified through thematic analysis and aligned with Skytrax evaluation criteria be prioritized to enhance Singapore Airlines' service offerings and sustain its industry-leading position?
4. How can trends in the prevalence of main topics within positive and negative sentiments be used to forecast changes in customer feedback and anticipate future service priorities?

Literature Overview

Sentiment Analysis with DistilBERT

Understanding customer feedback through sentiment analysis is critical for identifying factors that drive satisfaction and dissatisfaction in service industries like airlines (Kwon et al., 2021).

Traditional sentiment analysis tools, such as VADER and TextBlob, are widely used for their simplicity and ease of implementation. However, both have significant limitations. VADER performs well for short and straightforward texts but struggles with complex sentence structures and sarcasm, while TextBlob often fails to capture emotive nuances in detailed and context-rich reviews (Al-Adaileh et al., 2024).

Advanced transformer-based models, such as BERT (Bidirectional Encoder Representations from Transformers), have been developed to address these shortcomings by providing deep contextual understanding. However, BERT's computational intensity makes it less suitable for large datasets. DistilBERT, a lightweight version of BERT, retains 97% of BERT's performance while being 60% faster and using 40% fewer resources (Sanh et al., 2019). This efficiency makes DistilBERT particularly suitable for analyzing extensive datasets like the 10,000 TripAdvisor reviews in this study, allowing for accurate sentiment classification into positive and negative categories. The binary classification approach highlights actionable insights by focusing on areas where service improvements can be prioritized.

Forecasting Trends in Customer Sentiments and Themes

Time series forecasting plays a critical role in anticipating trends in customer feedback and planning strategic interventions. Various forecasting methods are available, each with its strengths and limitations. SARIMA (Seasonal Autoregressive Integrated Moving Average) is a widely recognized statistical model for forecasting seasonal time series data. While effective for datasets with clear seasonal and trend patterns, SARIMA requires meticulous parameter tuning and struggles with datasets exhibiting abrupt changes, such as those influenced by the COVID-19 pandemic (Hyndman & Athanasopoulos, 2018). Similarly, the ETS (Error, Trend, Seasonality) model provides an interpretable approach by directly modeling error, trend, and seasonality components. Although simpler to implement, ETS is less effective for datasets with complex, non-linear patterns (Taylor & Letham, 2018).

In contrast to these traditional models, LSTM (Long Short-Term Memory) networks and Facebook Prophet provide more flexible and adaptive forecasting solutions. LSTM networks, a type of recurrent neural network, excel in capturing non-linear dependencies and long-term trends, making them suitable for dynamic datasets (Hochreiter & Schmidhuber, 1997). Facebook Prophet is a user-friendly tool designed for forecasting time series data with irregularities, seasonal effects, and missing values. Its ease of implementation and interpretability make it particularly advantageous for analyzing customer sentiment and thematic trends (Taylor & Letham, 2018).

This study employs LSTM and Facebook Prophet to forecast the prevalence of main topics associated with customer sentiments. These methods were selected for their ability to adapt to non-linear and irregular data patterns, providing insights into emerging trends in customer feedback and service priorities. Their complementary strengths make them a more suitable choice than SARIMA or ETS for analyzing data influenced by structural breaks and disruptions like the pandemic.

Thematic Analysis with BERTopic and Zero-Shot Classification

Thematic analysis is an essential tool for uncovering deeper insights into customer experiences. Traditional methods, such as Latent Dirichlet Allocation (LDA), are widely used for topic modeling but often fail to capture contextual nuances, especially in reviews containing diverse themes (Grootendorst, 2022). BERTopic addresses these limitations by leveraging BERT embeddings to identify contextually relevant themes with high accuracy. Its ability to dynamically adjust the number of topics without requiring pre-defined parameters makes it ideal for analyzing diverse and unstructured user-generated content.

Despite its strengths, BERTopic faces limitations when reviews contain multiple overlapping themes, such as food, seating, and service quality, within a single review. To overcome this, Zero-Shot Classification is employed as a complementary approach. Zero-Shot Classification uses large transformer models to assign the most relevant topic to reviews with overlapping themes, ensuring a more granular understanding of customer feedback. This dual approach enables the mapping of identified themes to the Skytrax evaluation criteria, allowing for benchmarking and deeper insights into why Singapore Airlines consistently ranks among the top three airlines.

Thematic trends are further analyzed across three critical phases—pre-COVID, during COVID-19, and post-COVID—to explore how customer experiences and priorities have shifted. This comprehensive thematic analysis provides actionable insights into areas for service improvement and strategic planning.

\

Data Preparation with NLTK

Preprocessing is a critical step to ensure clean and consistent data for analysis. This study employs NLTK (Natural Language Toolkit) for tokenization, stop word removal, and text normalization (Bird, Klein, & Loper, 2009). These preprocessing steps minimize noise in the dataset, optimizing the performance of sentiment and thematic analysis models.

Proposed Methodology

Data Preprocessing

To ensure the quality and consistency of the dataset before analysis, a thorough data preprocessing step is implemented. The process involves cleaning the text data by removing irrelevant content, such as advertisements, and standardizing text formats. NLTK (Natural Language Toolkit) is employed for tokenizing the text, removing stopwords, and normalizing the data by converting it to lowercase and eliminating special characters. This preparation ensures that the data is free from noise and ready for accurate sentiment analysis and thematic exploration, thereby enhancing the reliability of the results.

Dataset Description

The dataset comprises 10,000 anonymized customer reviews from TripAdvisor, collected between 2018 and 2024. These reviews were obtained through a structured web scraping process designed to ensure the data's authenticity and relevance. Data verification included steps to remove duplicates and inconsistencies, while anonymization ensured that no personal identifiers were retained. This dataset provides valuable insights into customer sentiments related to service quality, especially in the context of changing expectations during and after the COVID-19 pandemic.

Sentiment Analysis Using DistilBERT

Sentiment analysis is performed using DistilBERT, a transformer-based model optimized for efficient text analysis while retaining deep contextual understanding. Each review is classified into binary sentiment categories: positive or negative. This binary classification is crucial for identifying actionable insights, as it focuses on highlighting areas for service improvement based on negative feedback while reinforcing positive attributes. This analysis is conducted across three distinct time periods: pre-COVID, during COVID-19, and post-COVID.

Visualization with Word Clouds

Following sentiment classification, Word Cloud visualizations are created to identify frequently mentioned keywords and phrases within each sentiment category (positive and negative). These visualizations offer an intuitive overview of the most common topics and concerns raised by customers, providing a foundation for deeper thematic analysis.

Forecasting Trends with LSTM and Facebook Prophet

To forecast future trends in customer feedback, this study employs LSTM (Long Short-Term Memory) networks and Facebook Prophet. LSTM networks are particularly effective for modeling complex, non-linear dependencies in time series data, while Facebook Prophet is well-suited for handling irregularities and seasonal variations. Both methods are trained and evaluated using metrics such as RMSE, MAE, and MAPE to determine their suitability for the dataset. This approach focuses on forecasting the prevalence of main topics associated with sentiments, enabling Singapore Airlines to anticipate shifts in customer priorities.

Thematic Analysis Using BERTopic and Zero-Shot Classification

To extract actionable insights into specific service attributes, thematic analysis is conducted using BERTopic and Zero-Shot Classification. BERTopic, which leverages BERT embeddings, identifies key themes and topics in customer reviews, offering a dynamic and context-aware approach to topic modeling. To provide an intuitive visualization of the relationships between topics, Intertopic Distance Maps are generated. These maps use a two-dimensional representation to display how closely topics are related to one another, helping to identify overarching themes and clusters of feedback.

Despite its strengths, BERTopic faces limitations when reviews contain multiple overlapping themes, such as food, seating, and service quality, within a single review. To overcome this, Zero-Shot Classification is used as a complementary tool to assign the most relevant topic to these reviews. This hybrid approach ensures a comprehensive understanding of customer feedback.

The identified themes are mapped to the Skytrax evaluation criteria to benchmark the findings against structured industry standards. Furthermore, thematic trends are analyzed across the three time periods to reveal how customer expectations and experiences have evolved over time.

Integration of Sentiment and Thematic Insights

The insights from sentiment analysis and thematic analysis are integrated to provide a comprehensive understanding of customer feedback. This integration highlights specific service attributes that drive satisfaction or dissatisfaction, offering actionable recommendations for service improvement and strategic planning.

Comparison with Skytrax Assessment

The results of the analysis are compared with the structured evaluation criteria of the Skytrax World Airline Awards. By aligning customer feedback with these criteria, the study validates customer perceptions against an established industry benchmark, offering a holistic view of service quality.

Limitations

The study acknowledges several limitations. Although DistilBERT effectively captures nuanced sentiment, it may face challenges in interpreting sarcasm or highly ambiguous text. Additionally, the reliance on TripAdvisor reviews introduces potential biases, as the platform often attracts extreme opinions. Future research could address these limitations by incorporating reviews from multiple platforms, such as Google Reviews or airline-specific feedback surveys, to enhance the comprehensiveness of the analysis.

Tools and Software

The study leverages Python as the primary programming language, utilizing the following libraries and tools:

1. Hugging Face Transformers for implementing DistilBERT in sentiment analysis.
2. BERTopic and Zero-Shot Classification for thematic analysis and topic modeling.
3. Intertopic Distance Maps for visualizing relationships between topics.
4. NLTK for data preprocessing tasks, including tokenization, stopword removal, and text normalization.
5. Facebook Prophet and TensorFlow/Keras for forecasting trends using LSTM and Prophet models.
6. Pandas and NumPy for data manipulation and analysis.
7. Matplotlib and Word Cloud for visualizing sentiment trends, thematic insights, and keyword frequencies.

Data Analysis Plan

1. Temporal Sentiment Analysis Using DistilBERT

To address Research Question 1, DistilBERT will be employed to classify reviews into binary sentiment categories: positive and negative. The analysis will track changes in sentiment across three distinct time periods: pre-COVID, during COVID-19, and post-COVID. By examining shifts in sentiment distribution over time, this step provides insights into how customer perceptions have evolved in response to significant events such as the COVID-19 pandemic. Key insights from this temporal sentiment analysis will highlight trends in satisfaction and dissatisfaction, offering a foundation for further exploration of underlying factors.

2. Thematic Analysis with BERTopic and Zero-Shot Classification

For Research Question 2, thematic analysis will be conducted using BERTopic to identify key themes and topics associated with positive and negative sentiments. The analysis will explore how these themes vary across the three time periods to reveal the evolving priorities and concerns of customers. For reviews containing multiple overlapping themes, Zero-Shot Classification will be employed to identify the most relevant topic, ensuring a more granular and accurate understanding of customer feedback. Additionally, Intertopic Distance Maps will be generated to visualize the relationships between topics, providing an overview of thematic clusters and their connections to sentiments.

3. Mapping Themes to Skytrax Evaluation Criteria

To address Research Question 3, the identified themes will be mapped to the structured evaluation criteria of the Skytrax World Airline Awards. This step aligns unstructured customer feedback with industry benchmarks, revealing areas where Singapore Airlines excels and opportunities for service enhancement. By prioritizing themes that strongly correlate with customer sentiments and aligning them with Skytrax standards, actionable recommendations for service improvement will be developed. This analysis focuses on key service attributes such as in-flight amenities, seating comfort, and customer service.

4. Forecasting Trends in Thematic Prevalence with LSTM and Facebook Prophet

For Research Question 4, forecasting will be conducted to predict trends in the prevalence of main topics associated with positive and negative sentiments. LSTM networks and Facebook

Prophet will be utilized to model these trends, leveraging their ability to handle non-linear dependencies and seasonal effects. Forecasting will provide insights into whether specific topics are likely to gain or lose prominence in customer reviews, helping Singapore Airlines anticipate shifts in customer priorities and adjust their strategies accordingly. Metrics such as RMSE, MAE, and MAPE will be used to evaluate the forecasting models and determine their suitability for the dataset.

5. Integration of Sentiment and Thematic Insights

The findings from sentiment analysis and thematic analysis will be integrated to provide a comprehensive understanding of customer feedback. This integration will reveal the specific service attributes that drive satisfaction or dissatisfaction, offering actionable recommendations for improvement. The insights will also inform long-term strategic planning to align services with evolving customer expectations.

6. Visualization and Reporting

To effectively communicate the results, Matplotlib and Word Cloud visualizations will be used to present sentiment trends, thematic insights, and forecasting outcomes. Intertopic Distance Maps will be included to visually represent thematic clusters, making it easier to identify key patterns and connections. These visualizations will ensure that findings are clearly presented to aid decision-making and strategic planning for service enhancements.

Potential Limitations

While this study employs a robust and comprehensive methodology, several limitations must be considered. The dataset, sourced exclusively from TripAdvisor, may introduce inherent bias as reviews on this platform often represent polarized opinions—either highly positive or highly negative experiences. This selection bias could lead to an underrepresentation of moderate customer experiences, potentially skewing sentiment analysis results. Although the use of sentiment-based segmentation mitigates some of this bias, it does not fully eliminate its impact on the analysis.

Sentiment analysis, conducted using DistilBERT, provides enhanced contextual understanding compared to traditional models. However, it may still struggle with subtle linguistic nuances such as sarcasm, irony, or mixed sentiments within a single review. These complexities can lead to occasional misclassification, particularly in reviews that convey conflicting emotions or ambiguous language. Despite the advanced capabilities of transformer-based models, the accurate interpretation of nuanced sentiments remains a challenge in natural language processing.

Thematic analysis, performed using BERTopic, excels at identifying contextually rich topics and themes but has limitations when dealing with reviews that are either extremely brief or overly detailed. Short reviews may lack sufficient context for theme identification, while highly detailed reviews may discuss multiple unrelated topics, fragmenting the analysis. To address this, Zero-Shot Classification is used as a complementary tool; however, this additional step may introduce its own complexities, such as reliance on predefined topic categories that might not fully align with user-generated content.

The study's temporal scope, spanning reviews from 2018 to 2024, offers valuable insights into sentiment trends, particularly in relation to the COVID-19 pandemic. However, this time frame inherently limits the analysis of long-term shifts or emerging trends beyond 2024. Forecasting trends using LSTM and Facebook Prophet is based on historical data, which may not account for unforeseen disruptions or shifts in customer behavior post-2024. Predictive models are inherently limited by the assumptions underlying the data, which may reduce their accuracy in rapidly changing contexts.

Integrating findings with the structured evaluation criteria of the Skytrax World Airline Awards adds value by aligning unstructured customer feedback with industry benchmarks. However, direct comparisons between unstructured reviews and formal survey-based evaluations can be challenging. User-generated reviews often lack the rigor and consistency of structured survey data, complicating efforts to draw precise one-to-one correlations between the two sources.

Finally, this study focuses exclusively on Singapore Airlines using TripAdvisor data, which limits the generalizability of its findings. Insights derived from this dataset may not extend to other airlines or broader customer segments without further validation. Expanding the analysis to include reviews from multiple platforms, such as Google Reviews or AirlineRatings, or incorporating data from different airlines, would improve the robustness and generalizability of the results. Additionally, exploring alternative modeling approaches or extending the temporal scope of the analysis would further enhance the depth and applicability of future research.

By acknowledging these limitations, the study provides a transparent basis for interpreting its findings and offers opportunities for future research to build on its methodology, expand its scope, and validate its results in broader contexts.

Ethical Considerations

This study adheres to rigorous ethical standards throughout the data management lifecycle, from collection to analysis and reporting. The dataset consists of 10,000 anonymized TripAdvisor reviews, collected systematically using Python-based web scraping tools for the period from 2018 to 2024. The data collection process ensured authenticity through strict validation procedures, including the removal of duplicates, irrelevant content, and any information that could compromise data integrity. Personal identifiers were anonymized to protect user privacy, fully aligning with ethical principles of data confidentiality and compliance with applicable regulations.

Particular attention is paid to mitigating biases inherent in user-generated content. Reviews on platforms such as TripAdvisor may reflect biases based on gender, nationality, or cultural background, potentially influencing sentiment analysis outcomes. To address this, the study employs a balanced analytical framework that emphasizes aggregate trends rather than individual reviews. While the dataset is exclusively sourced from TripAdvisor, the themes and sentiment patterns identified are cross-referenced with publicly available benchmarks, such as Skytrax evaluation criteria, to ensure a broader and more inclusive perspective. This step enhances the study's robustness without requiring additional data collection, thereby maintaining ethical integrity.

Sentiment analysis is conducted using DistilBERT, a transformer-based model optimized for efficient text analysis. The limitations of DistilBERT, particularly in detecting subtle nuances such as sarcasm, irony, or mixed sentiments, are carefully acknowledged. The analysis focuses on aggregated trends rather than detailed interpretations of individual reviews, reducing the potential for overgeneralization or misinterpretation. This ensures that insights remain aligned with the study’s objectives while respecting the complexities of human language.

Thematic analysis is performed with BERTopic and Zero-Shot Classification, and ethical considerations extend to ensuring that these models do not unintentionally perpetuate biases present in the data. Themes are derived and analyzed transparently, with results presented in an aggregated form to protect the anonymity of individual reviewers. Intertopic Distance Maps and other visualizations are used to highlight general trends and relationships without exposing specific review content.

The reporting process emphasizes transparency and integrity. Potential biases and limitations inherent in the dataset and analytical methods are clearly disclosed, enabling an objective interpretation of the findings. All insights are used solely for academic purposes, with actionable recommendations for Singapore Airlines derived from anonymized and aggregated data. The study avoids overinterpretation and ensures that results are framed within the scope of the research objectives.

Upon completion of the study, the dataset will be securely destroyed in accordance with stringent data retention policies. This ensures compliance with ethical standards, prevents potential misuse of the data, and reinforces academic integrity. By prioritizing transparency, confidentiality, and ethical rigor, this study safeguards the interests of all stakeholders while contributing meaningful insights to the field of customer experience analysis.

Project Plan and Timeline

Topic	Week										
	1	2	3	4	5	6	7	8	9	10	11
Write the Project Proposal	■										
Data Collection and Preprocessing		■									
Sentiment Analysis with DistilBERT			■	■							
Temporal Sentiment Analysis					■						
Thematic Analysis with BERTopic						■	■				
Forecasting Thematic Trends								■			
Insights and Recommendations									■		
Write Business Analytics Report										■	
Final Report Submission											■

Week	Date	Content
1	21 Oct - 27 Oct	Drafting and submitting the project proposal, including refining research questions, methodology, and expected outcomes.
2	28 Oct - 3 Nov	Collecting and preprocessing the dataset (10,000 TripAdvisor reviews). This involves cleaning, tokenizing text, and preparing data using NLTK to ensure quality and consistency for further analysis.
3	4 Nov - 17 Nov	Conducting Sentiment Analysis using DistilBERT to classify reviews into positive and negative categories. This forms the basis for tracking sentiment shifts over the COVID-19 timeline.
4		
5	18 Nov - 24 Nov	Performing Temporal Sentiment Analysis to examine shifts in sentiment across the pre-COVID, during COVID-19, and post-COVID periods. Identifying frequently mentioned keywords within each sentiment category using NLTK for sentiment-based segmentation.
6	25 Nov - 8 Dec	mplementing Thematic Analysis with BERTopic and Zero-Shot Classification to identify key themes and service attributes. Generating Intertopic Distance Maps to visualize relationships between themes. This analysis addresses Research Question 2.
7		
8	9 Dec - 15 Dec	Conducting Forecasting of Future Thematic Trends using LSTM and Facebook Prophet. Predicting shifts in the prevalence of main topics within positive and negative sentiments to anticipate changes in customer priorities. This step addresses Research Question 4.
9	16 Dec - 22 Dec	Synthesizing insights from sentiment analysis, thematic analysis, and forecasting to develop actionable recommendations for service enhancement. Addressing Research Question 3 by mapping themes to Skytrax evaluation criteria.
10	23 Dec - 29 Dec	Writing the Business Analytics Report. Integrating findings from the analyses to create a comprehensive report that includes methodologies, key findings, and actionable insights.
11	30 Dec - 3 Jan	Reviewing, refining, and finalizing the report. Preparing the Executive Summary and ensuring that all deliverables are completed for submission by the deadline of 3 January 2025 (12 PM).

Executive Summary Plan

The Executive Summary provides a concise overview of the project, presenting essential details about the research objectives, methodologies, findings, and actionable recommendations. This section is designed to offer business stakeholders and academic readers a clear understanding of the study's significance and its outcomes.

Introduction

The introduction begins by outlining the importance of customer satisfaction for Singapore Airlines, emphasizing its consistent top-three ranking in the Skytrax World Airline Awards. It highlights the study's aim to analyze customer feedback from TripAdvisor reviews and align findings with Skytrax evaluation criteria to enhance service quality and sustain competitive leadership. The research questions are then summarized, addressing how customer sentiment evolves over time, the themes associated with positive and negative sentiments, the alignment of these themes with Skytrax criteria, and the forecasting of thematic trends.

Methodology

The methodology section describes the dataset of 10,000 anonymized TripAdvisor reviews collected between 2018 and 2024, emphasizing its relevance in capturing customer expectations before, during, and after the COVID-19 pandemic. It explains the analytical approach, which combines advanced tools such as DistilBERT for sentiment analysis, BERTopic and Zero-Shot Classification for thematic analysis, and LSTM and Facebook Prophet for forecasting. The integration of these techniques enables an in-depth analysis of sentiment trends, theme identification, and the prediction of future customer priorities, supported by effective visualizations.

Key Findings

Key findings include the observed sentiment trends, highlighting the proportions of positive and negative sentiments and their evolution across the pre-COVID, during COVID-19, and post-COVID periods. Major shifts in customer feedback are analyzed to reveal patterns over time. Thematic analysis identifies recurring topics such as in-flight service, seating comfort, and amenities, while exploring their correlation with positive and negative sentiments. These themes are then compared with the structured evaluation criteria of Skytrax, offering a benchmark to understand Singapore Airlines' strengths and areas for improvement.

Forecasting Insights

The forecasting insights summarize predicted trends, focusing on the anticipated changes in the prevalence of main topics within positive and negative sentiments. These predictions provide a clearer understanding of emerging customer priorities and potential adjustments Singapore Airlines might consider to address future expectations.

Actionable Recommendations

Actionable recommendations are derived from the sentiment and thematic analysis, emphasizing improvements in specific service attributes such as customer service, in-flight amenities, and seating arrangements. These recommendations are carefully aligned with Skytrax criteria to ensure their relevance and potential for meaningful impact on customer satisfaction.

Conclusion

The conclusion highlights the significance of the study's findings for Singapore Airlines, emphasizing their relevance in maintaining leadership and enhancing customer satisfaction. It underscores the value of aligning unstructured customer feedback with structured industry benchmarks. Finally, it suggests future research directions, including the potential for expanding the dataset to incorporate reviews from additional platforms and exploring alternative forecasting models to enhance predictive accuracy.

Reference Lists

Al-Adaileh, A., Al-Kfairy, M., Tubishat, M., & Alfandi, O. (2024). A sentiment analysis approach for understanding users' perception of metaverse marketplace. Intelligent Systems With Applications, 22, 200362. https://doi.org/10.1016/j.iswa.2024.200362

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python (1st edition). O'Reilly.

Felix, C., Franconeri, S., & Bertini, E. (2017). Taking Word Clouds Apart: An empirical investigation of the design space for keyword summaries. IEEE Transactions on Visualization and Computer Graphics, 24(1), 657–666. https://doi.org/10.1109/tvcg.2017.2746018

Grootendorst, M. (2022). BERTopic: Leveraging BERT embeddings for topic modeling. https://arxiv.org/pdf/2203.05794

Hutto, C.J. & Gilbert, E. (2015). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014. https://www.researchgate.net/publication/275828927_VADER_A_Parsimonious_Rule-based_Model_for_Sentiment_Analysis_of_Social_Media_Text

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting principles and practice.

Ketchen, D. J., & Shook, C. L. (1996). The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. Strategic Management Journal, 17(6), 441–458. http://www.jstor.org/stable/2486927

Lohmann, G., & Koo, T. T. R. (2013). The impact of airline service quality on passengers' satisfaction and choice behavior. Journal of Air Transport Management, 34, 28-31. https://doi.org/10.1016/j.jairtraman.2013.07.001

Park, S., Cho, J., Park, K., & Shin, H. (2021). Customer sentiment analysis with more sensibility. Engineering Applications of Artificial Intelligence, 104, 104356. https://doi.org/10.1016/j.engappai.2021.104356

Wang, K. (2023). Introduction to clustering-based customer segmentation. Medium. https://medium.com/data-science-at-microsoft/introduction-to-clustering-based-customer-segmentation-2fac61e80100

Appendix C: Python code

This section is contain the following analysis :

Data Collection and Preprocessing (Week 2)

1. Load and Explore the Dataset
2. Data Cleaning
3. Text Preprocessing with NLTK
4. Exploratory Data Analysis (EDA)

Week2 : Data Collection and Preprocessing

1. Load and Explore the Dataset

[2]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[3]:
```python
# Load the dataset
df = pd.read_csv('singapore_airlines_reviews.csv', encoding='utf-8')

# Get the shape of the dataset (number of rows and columns)
print(f"Dataset contains {df.shape[0]} rows and {df.shape[1]} columns.")

# Get information about data types and missing values
df.info()

# Check for missing values in each column
missing_values = df.isnull().sum()
print("Missing values per column:")
print(missing_values)

# Summary statistics of numerical columns (if any)
df.describe()
```

```
Dataset contains 10000 rows and 7 columns.
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 7 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   published_date      10000 non-null  object
 1   published_platform  10000 non-null  object
 2   rating              10000 non-null  int64
 3   type                10000 non-null  object
 4   text                10000 non-null  object
 5   title               9999 non-null   object
 6   helpful_votes       10000 non-null  int64
dtypes: int64(2), object(5)
memory usage: 547.0+ KB
Missing values per column:
published_date        0
published_platform    0
rating                0
type                  0
text                  0
title                 1
helpful_votes         0
dtype: int64
```

[3]:

	rating	helpful_votes
count	10000.000000	10000.000000
mean	4.015800	1.275200
std	1.346006	2.721618
min	1.000000	0.000000
25%	3.000000	0.000000
50%	5.000000	1.000000
75%	5.000000	2.000000
max	5.000000	158.000000

2. Data Cleaning

[4]:
```python
# Check for duplicates
duplicates = df.duplicated().sum()
print(f"Number of duplicate rows: {duplicates}")

# Drop duplicates if found
df.drop_duplicates(inplace=True)
print(f"Dataset after removing duplicates: {df.shape[0]} rows")
```

```
Number of duplicate rows: 0
Dataset after removing duplicates: 10000 rows
```

[5]:
```python
# Display column names
print("Column names:")
print(df.columns)

# Check for unique values in each column (useful if there are categorical columns)
for col in df.columns:
    print(f"Unique values in {col}: {df[col].nunique()}")
```

```
Column names:
Index(['published_date', 'published_platform', 'rating', 'type', 'text',
       'title', 'helpful_votes'],
      dtype='object')
Unique values in published_date: 9997
Unique values in published_platform: 2
Unique values in rating: 5
Unique values in type: 1
Unique values in text: 10000
Unique values in title: 8476
Unique values in helpful_votes: 27
```

3. Text Preprocessing with NLTK

```
pip install nltk
```

```
Requirement already satisfied: nltk in ./anaconda3/lib/python3.11/site-packages (3.8.1)
Requirement already satisfied: click in ./anaconda3/lib/python3.11/site-packages (from nltk) (
Requirement already satisfied: joblib in ./anaconda3/lib/python3.11/site-packages (from nltk)
Requirement already satisfied: regex>=2021.8.3 in ./anaconda3/lib/python3.11/site-packages (fr
Requirement already satisfied: tqdm in ./anaconda3/lib/python3.11/site-packages (from nltk) (4
Note: you may need to restart the kernel to use updated packages.
```

```
import nltk
import re
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer

# Download required NLTK datasets (run only once)
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('omw-1.4')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     /Users/natamontosawat/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]     /Users/natamontosawat/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data]     /Users/natamontosawat/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package omw-1.4 to
[nltk_data]     /Users/natamontosawat/nltk_data...
[nltk_data]   Package omw-1.4 is already up-to-date!
```

```
True
```

```
# Initialize the lemmatizer
lemmatizer = WordNetLemmatizer()

# Load English stopwords
stop_words = set(stopwords.words('english'))

def clean_text(text):
    # Convert to lowercase
    text = text.lower()

    # Remove special characters, numbers, and punctuation
    text = re.sub(r'[^a-z\s]', '', text)

    # Tokenize the text
    tokens = word_tokenize(text)

    # Remove stopwords and lemmatize the words
    cleaned_tokens = [lemmatizer.lemmatize(word) for word in tokens if word not in stop_words]

    # Join the cleaned tokens back into a single string
    cleaned_text = ' '.join(cleaned_tokens)

    return cleaned_text
```

```
# Apply the cleaning function to the 'review_text' column
df['cleaned_review'] = df['text'].apply(lambda x: clean_text(str(x)))

# Display the cleaned data
df[['text', 'cleaned_review']].head()
```

	text	cleaned_review
0	We used this airline to go from Singapore to L...	used airline go singapore london heathrow issu...
1	The service on Singapore Airlines Suites Class...	service singapore airline suite class nothing ...
2	Booked, paid and received email confirmation f...	booked paid received email confirmation extra ...
3	Best airline in the world, seats, food, servic...	best airline world seat food service brilliant...
4	Premium Economy Seating on Singapore Airlines ...	premium economy seating singapore airline narr...


```
# Save the cleaned dataset to a new CSV file
df.to_csv('cleaned_singapore_airlines_reviews.csv', index=False, encoding='utf-8')
print("Cleaned dataset saved as 'cleaned_singapore_airlines_reviews.csv'.")
```

Week 3-4 Sentiment Analysis

1. Perform Sentiment Classification

1.1 Load the VADER and DistilBERT Models and Tokenizer

]:
```
from transformers import DistilBertTokenizer

# lode tokenizer from pretrained model
tokenizer = DistilBertTokenizer.from_pretrained('distilbert-base-uncased')
```

]:
```
# Import libraries
from transformers import DistilBertTokenizer
```

]:
```
from transformers import DistilBertTokenizer

def preprocess_text(text, tokenizer):
    inputs = tokenizer(
        text,
        return_tensors='pt',
        max_length=128,
        truncation=True,
        padding='max_length'
    )
    return inputs
```

]:
```
# Load the cleaned dataset
import pandas as pd

df = pd.read_csv('cleaned_singapore_airlines_reviews.csv')

df['inputs'] = df['cleaned_review'].apply(lambda x: preprocess_text(x, tokenizer))
```

]:
```
# Import required libraries
import torch
from transformers import DistilBertTokenizer

# Check device
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
print(f"Using device: {device}")

# Load tokenizer
tokenizer = DistilBertTokenizer.from_pretrained('distilbert-base-uncased')

# Define preprocessing function
def preprocess_text(text):
    inputs = tokenizer(
        text,
        return_tensors='pt',
        max_length=128,
        truncation=True,
        padding='max_length'
    )
    inputs = {key: value.to(device) for key, value in inputs.items()}
    return inputs

# Apply preprocessing
df['inputs'] = df['cleaned_review'].apply(preprocess_text)
```

```
Using device: cpu
```

1.2 Preprocess Reviews

```python
import pandas as pd
# Load the cleaned dataset
df = pd.read_csv('cleaned_singapore_airlines_reviews.csv')

# Function to preprocess the text for DistilBERT
def preprocess_text(text):
    # Tokenize the input text and convert to tensors
    inputs = tokenizer(text, return_tensors='pt', max_length=128, truncation=True, padding='max_length')
    inputs = {key: value.to(device) for key, value in inputs.items()}
    return inputs

# Apply preprocessing to all reviews in your cleaned dataset
df['inputs'] = df['cleaned_review'].apply(preprocess_text)
```

```python
from transformers import DistilBertForSequenceClassification

# Load the pre-trained model
model = DistilBertForSequenceClassification.from_pretrained(
    'distilbert-base-uncased-finetuned-sst-2-english'
)
model.to(device)  # Move the model to the specified device (CPU or GPU)
```

```
DistilBertForSequenceClassification(
  (distilbert): DistilBertModel(
    (embeddings): Embeddings(
      (word_embeddings): Embedding(30522, 768, padding_idx=0)
      (position_embeddings): Embedding(512, 768)
      (LayerNorm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)
      (dropout): Dropout(p=0.1, inplace=False)
    )
    (transformer): Transformer(
      (layer): ModuleList(
        (0-5): 6 x TransformerBlock(
          (attention): DistilBertSdpaAttention(
            (dropout): Dropout(p=0.1, inplace=False)
            (q_lin): Linear(in_features=768, out_features=768, bias=True)
            (k_lin): Linear(in_features=768, out_features=768, bias=True)
            (v_lin): Linear(in_features=768, out_features=768, bias=True)
            (out_lin): Linear(in_features=768, out_features=768, bias=True)
          )
          (sa_layer_norm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)
```

```python
import warnings

warnings.filterwarnings("ignore", category=FutureWarning)
```

1.3 Perform Sentiment Classification with DistilBERT

```python
from transformers import DistilBertTokenizer, DistilBertForSequenceClassification
import torch
import pandas as pd
import torch.nn.functional as F  # For softmax

# Define device
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")

# Load tokenizer
tokenizer = DistilBertTokenizer.from_pretrained('distilbert-base-uncased')

# Load model
model = DistilBertForSequenceClassification.from_pretrained(
    'distilbert-base-uncased-finetuned-sst-2-english'
)
model.to(device)

# Define classify_sentiment function
def classify_sentiment_and_score(review):
    """
    Classify sentiment and calculate compound score for a given review.

    Parameters:
        review (str): The review text to analyze.

    Returns:
        tuple: Sentiment label (positive/negative) and Compound Score.
    """
    # Tokenize and prepare input
    inputs = tokenizer(review, return_tensors="pt", truncation=True, padding=True).to(device)

    with torch.no_grad():
        outputs = model(**inputs)  # Get logits from the model
        logits = outputs.logits  # Extract raw predictions

        # Apply softmax to calculate probabilities
        probabilities = F.softmax(logits, dim=1).squeeze()

        # Calculate Compound Score
        compound_score = (probabilities[1] - probabilities[0]).item()

        # Determine sentiment label
        sentiment_map = {0: 'negative', 1: 'positive'}
        predicted_class = torch.argmax(logits, dim=1).item()
        sentiment_label = sentiment_map.get(predicted_class, 'neutral')

        return sentiment_label, compound_score

# Load DataFrame
df = pd.read_csv('cleaned_singapore_airlines_reviews.csv')

# Apply the function to each review and split the results into two columns
df[['sentiment', 'compound_score']] = df['cleaned_review'].apply(
    lambda x: pd.Series(classify_sentiment_and_score(x))
)

# Save the updated DataFrame
df.to_csv('classified_reviews_with_sentiment_and_compound.csv', index=False)
print("Dataset saved as 'classified_reviews_with_sentiment_and_compound.csv'.")
```

```
Dataset saved as 'classified_reviews_with_sentiment_and_compound.csv'.
```

1.4 Perform Sentiment Classification with VADER

```
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
import pandas as pd

# Initialize the VADER sentiment analyzer
analyzer = SentimentIntensityAnalyzer()

# Function to classify sentiment using VADER
def classify_sentiment_vader(text):
    """
    Classify sentiment and calculate compound score using VADER.

    Parameters:
        text (str): The review text to analyze.

    Returns:
        tuple: Sentiment label (positive/negative) and Compound Score.
    """
    # Get sentiment scores from VADER
    scores = analyzer.polarity_scores(text)

    # Determine sentiment based on compound score
    compound_score = scores['compound']
    if compound_score >= 0.00:
        sentiment_label = 'positive'
    elif compound_score < -0.00:
        sentiment_label = 'negative'


    return sentiment_label, compound_score

# Load the cleaned dataset
df = pd.read_csv('cleaned_singapore_airlines_reviews.csv')

# Apply the VADER sentiment analysis function to each review
df[['vader_sentiment', 'vader_compound_score']] = df['cleaned_review'].apply(
    lambda x: pd.Series(classify_sentiment_vader(x))
)

# Save the updated DataFrame with VADER results
df.to_csv('vader_sentiment_results.csv', index=False)
print("Dataset saved as 'vader_sentiment_results.csv'.")
```

```
Dataset saved as 'vader_sentiment_results.csv'.
```

2. Comparison of Model Results

```
import pandas as pd

# Load datasets
df = pd.read_csv('classified_reviews_with_sentiment_and_compound.csv')  # DistilBERT results
vader_df = pd.read_csv('vader_sentiment_results.csv')  # VADER results

# Merge DistilBERT and VADER results
df['vader_sentiment'] = vader_df['vader_sentiment']

# Count and percentage for DistilBERT
distilbert_counts = df['sentiment'].value_counts()
distilbert_percentages = distilbert_counts / len(df) * 100

# Count and percentage for VADER
vader_counts = df['vader_sentiment'].value_counts()
vader_percentages = vader_counts / len(df) * 100

# Display results
print("DistilBERT Sentiment Counts and Percentages:")
print(distilbert_counts)
print(distilbert_percentages)

print("\nVADER Sentiment Counts and Percentages:")
print(vader_counts)
print(vader_percentages)
```

```
DistilBERT Sentiment Counts and Percentages:
```

3. Evaluate Sentiment Classification

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, roc_auc_score

# Load the processed dataset with both DistilBERT and VADER sentiment results
df = pd.read_csv('classified_reviews_with_sentiment_and_compound.csv')  # DistilBERT
vader_df = pd.read_csv('vader_sentiment_results.csv')  # VADER
df['vader_sentiment'] = vader_df['vader_sentiment']

# Map sentiments to numeric labels (1 for positive, 0 for negative)
df['distilbert_numeric'] = df['sentiment'].map({'positive': 1, 'negative': 0})
df['vader_numeric'] = df['vader_sentiment'].map({'positive': 1, 'negative': 0})

# Drop rows with NaN values in 'distilbert_numeric' or 'vader_numeric'
df = df.dropna(subset=['distilbert_numeric', 'vader_numeric'])

# Prepare a surrogate ground truth using DistilBERT predictions
X = df['cleaned_review']
y = df['distilbert_numeric']  # Surrogate label based on DistilBERT

# Split data into training and testing sets (80/20 split)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Convert text data into TF-IDF features
tfidf = TfidfVectorizer(max_features=5000)
X_train_tfidf = tfidf.fit_transform(X_train)
X_test_tfidf = tfidf.transform(X_test)

# Train a Random Forest Classifier as a surrogate evaluation model
clf = RandomForestClassifier(random_state=42)
clf.fit(X_train_tfidf, y_train)

# Evaluate DistilBERT Sentiment Results
y_pred_distilbert = clf.predict(tfidf.transform(df['cleaned_review']))
y_pred_proba_distilbert = clf.predict_proba(tfidf.transform(df['cleaned_review']))[:, 1]

distilbert_report = classification_report(y, y_pred_distilbert, output_dict=True)
roc_auc_distilbert = roc_auc_score(y, y_pred_proba_distilbert)

print("DistilBERT Evaluation (Surrogate Model):")
print(classification_report(y, y_pred_distilbert))
print(f"ROC-AUC Score: {roc_auc_distilbert}")

# Evaluate VADER Sentiment Results
y = df['vader_numeric']  # Surrogate label based on VADER
y_pred_vader = clf.predict(tfidf.transform(df['cleaned_review']))
y_pred_proba_vader = clf.predict_proba(tfidf.transform(df['cleaned_review']))[:, 1]

vader_report = classification_report(y, y_pred_vader, output_dict=True)
roc_auc_vader = roc_auc_score(y, y_pred_proba_vader)

print("\nVADER Evaluation (Surrogate Model):")
print(classification_report(y, y_pred_vader))
print(f"ROC-AUC Score: {roc_auc_vader}")

# Save Reports to CSV
pd.DataFrame(distilbert_report).transpose().to_csv("distilbert_surrogate_report.csv")
pd.DataFrame(vader_report).transpose().to_csv("vader_surrogate_report.csv")

# Summarized Comparison
print("\nSummary Comparison:")
print(f"DistilBERT ROC-AUC: {roc_auc_distilbert}")
print(f"VADER ROC-AUC: {roc_auc_vader}")
```

```
DistilBERT Evaluation (Surrogate Model):
```

1. Thematic analysis Overview

1.1 Using BERTopic

$]:

```
import pandas as pd
from bertopic import BERTopic
from sentence_transformers import SentenceTransformer

# Load the original dataset with reviews, sentiment, and published_date
df = pd.read_csv('simplified_reviews_with_sentiment.csv')

# Ensure the 'published_date' is in datetime format
df['published_date'] = pd.to_datetime(df['published_date'], errors='coerce')
df = df.dropna(subset=['published_date'])

# Load the pre-trained model for embeddings
model = SentenceTransformer('all-MiniLM-L6-v2')

# Generate embeddings for the reviews
embeddings = model.encode(df['cleaned_review'].to_list(), show_progress_bar=True)

# Fit the BERTopic model and get topics for each review
topic_model = BERTopic()
topics, _ = topic_model.fit_transform(df['cleaned_review'].to_list(), embeddings)

# Add the topic IDs back to the original DataFrame
df['topic'] = topics

# Get topic information
topic_info = topic_model.get_topic_info()

# Map topic labels and keywords
def get_topic_details(topic_id):
    if topic_id == -1:  # Handle outliers
        return "Outlier", ""
    else:
        # Fetch topic information
        topic_row = topic_info[topic_info['Topic'] == topic_id]
        if topic_row.empty:  # If topic ID does not exist in topic_info
            return "Unknown", "N/A"
        else:
            topic_label = topic_row['Name'].values[0]
            keywords = ", ".join([kw[0] for kw in topic_model.get_topic(topic_id)])
            return topic_label, keywords

# Apply the mapping to add topic labels and keywords
df[['topic_label', 'topic_keywords']] = df['topic'].apply(
    lambda x: pd.Series(get_topic_details(x))
)

# Save the updated DataFrame with topics, labels, and keywords
df.to_csv('thematic_analysis_per_review_with_labels.csv', index=False)
print("Thematic analysis with topics, labels, and keywords saved successfully!")
```

```
Batches:   0%|          | 0/313 [00:00<?, ?it/s]
```

:

```
# function for Top 10 Topics
def print_top_topics(topic_model, reviews, sentiment):
    print(f"\nTop 10 Topics for {sentiment.capitalize()} Reviews:")
    topics, _ = topic_model.fit_transform(reviews)
    topic_info = topic_model.get_topic_info().head(11)
    for index, row in topic_info.iterrows():
        print(f"Topic {row['Topic']}: {row['Name']} (Count: {row['Count']})")

import pandas as pd
from bertopic import BERTopic
from sentence_transformers import SentenceTransformer

# Load Model for embeddings
model = SentenceTransformer('all-MiniLM-L6-v2')

# Load Dataset
df = pd.read_csv('thematic_analysis_per_review_with_labels.csv')

# Separate Positive and Negative Reviews
positive_reviews = df[df['sentiment'] == 'positive']['cleaned_review']
negative_reviews = df[df['sentiment'] == 'negative']['cleaned_review']

# Create embeddings
positive_embeddings = model.encode(positive_reviews.to_list(), show_progress_bar=True)
negative_embeddings = model.encode(negative_reviews.to_list(), show_progress_bar=True)

# Analyst with BERTopic for Positive Reviews
positive_topic_model = BERTopic()
positive_topic_model.fit(positive_reviews.to_list(), positive_embeddings)

# Analyst with BERTopic for Negative Reviews
negative_topic_model = BERTopic()
negative_topic_model.fit(negative_reviews.to_list(), negative_embeddings)

# print Top 5 Topics for Positive Reviews
print_top_topics(positive_topic_model, positive_reviews, 'positive')

# print Top 5 Topics for Negative Reviews
print_top_topics(negative_topic_model, negative_reviews, 'negative')
```

```
Batches:   0%|          | 0/176 [00:00<?, ?it/s]
```

1.2 Using Zero-Shot Classification for Topic -1

by Mapping with Skytrax Criteria

```
]: from transformers import pipeline
import pandas as pd
from tqdm import tqdm
import logging

# Set up logging for errors
logging.basicConfig(filename="error_log_zero_shot.txt", level=logging.ERROR, format="%(asctime)s - %(message)s")

# Load the dataset
df = pd.read_csv('thematic_analysis_per_review_with_labels.csv')
topic_minus_one_df = df[df['topic'] == -1].reset_index(drop=True)

# Initialize Zero-Shot Classification pipeline
classifier = pipeline("zero-shot-classification", model="facebook/bart-large-mnli", multi_label=True)

# Define candidate labels
candidate_labels = [
    "Boarding assistance",
    "Service friendliness / hospitality",
    "Service attentiveness / efficiency",
    "Tickets and refunds",
    "Assisting families",
    "Online booking and check-in services",
    "Baggage delivery",
    "Seat comfort",
    "Cleanliness",
    "Meal service efficiency",
    "Entertainment",
    "Airline Lounge : product facilities"
]

# Add empty columns for results
topic_minus_one_df['predicted_topic'] = None
topic_minus_one_df['confidence'] = None

# Process each review with tqdm (Progress Bar)
for idx, review in tqdm(topic_minus_one_df['cleaned_review'].items(), desc="Classifying topics",
                        total=len(topic_minus_one_df)):
    try:
        # Skip empty reviews
        if not isinstance(review, str) or review.strip() == "":
            topic_minus_one_df.at[idx, 'predicted_topic'] = "General Topic"
            topic_minus_one_df.at[idx, 'confidence'] = 0.0
            continue

        # Run Zero-Shot Classification
        result = classifier(review, candidate_labels)
        topic_minus_one_df.at[idx, 'predicted_topic'] = result['labels'][0]
        topic_minus_one_df.at[idx, 'confidence'] = result['scores'][0]

    except Exception as e:
        # Log error and assign default values
        logging.error(f"Error processing index {idx}: {e}")
        topic_minus_one_df.at[idx, 'predicted_topic'] = "General Topic"
        topic_minus_one_df.at[idx, 'confidence'] = 0.0

# Save the fixed dataset
topic_minus_one_df.to_csv('zero_shot_classified_topic_minus_one_fixed.csv', index=False)
print("\nClassified topics saved successfully as 'zero_shot_classified_topic_minus_one_fixed.csv'")
```

```
Classifying topics: 100%|██████████| 5696/5696 [1:14:40<00:00,  1.27it/s]

Classified topics saved successfully as 'zero_shot_classified_topic_minus_one_fixed.csv'
```

```python
import pandas as pd

# Load the dataset
df = pd.read_csv('zero_shot_classified_topic_minus_one_fixed.csv')

# The number of topics was chosen.
print(df['predicted_topic'].value_counts())
```

```
predicted_topic
Service attentiveness / efficiency      1945
Service friendliness / hospitality      1442
Seat comfort                            1099
Meal service efficiency                  304
Tickets and refunds                      216
Entertainment                            199
Boarding assistance                      180
Online booking and check-in services     109
Assisting families                        86
Airline Lounge : product facilities       56
Cleanliness                               35
Baggage delivery                          25
Name: count, dtype: int64
```

```python
# Number of low confidence reviews(less than 0.6)
low_confidence_reviews = df[df['confidence'] < 0.6]
print(f"Number of low confidence reviews: {len(low_confidence_reviews)}")
```

```
Number of low confidence reviews: 188
```

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
file_path = 'zero_shot_classified_topic_minus_one_fixed.csv'
df = pd.read_csv(file_path)

# Get the count of predicted topics
topic_counts = df['predicted_topic'].value_counts()

# Create the bar chart with topic counts
plt.figure(figsize=(12, 8))
bars = plt.bar(topic_counts.index, topic_counts.values, color='darkblue')

# Add the count labels on top of each bar
for bar in bars:
    plt.text(bar.get_x() + bar.get_width() / 2 - 0.1, bar.get_height() + 10,
             str(bar.get_height()), ha='center', va='bottom')

# Customize the chart
plt.title('Topic Distribution of Topics-1', fontsize=16)
plt.xlabel('Predicted Topics', fontsize=14)
plt.ylabel('Count', fontsize=14)
plt.xticks(rotation=45, ha='right', fontsize=12)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()

# Display the plot
plt.show()
```

1.3 Mapping all topics with Skytrax Criteria

```python
import pandas as pd

# Load datasets
bertopic_df = pd.read_csv('thematic_analysis_per_review_with_labels.csv')
zero_shot_df = pd.read_csv('zero_shot_classified_topic_minus_one_fixed.csv')

# Step 1: Filter reviews where topic != -1
bertopic_mapped_df = bertopic_df[bertopic_df['topic'] != -1].copy()

# Define target topics with associated keywords
target_topics = {
    "Meal service efficiency": ["meal", "food", "drink", "dinner", "service", "meals", "breakfast"],
    "Seat comfort": ["seat", "legroom", "spacious", "comfort", "bed", "flat"],
    "Cleanliness": ["cleanliness", "tidy", "hygiene"],
    "Service friendliness / hospitality": ["service", "friendly", "hospitality", "kind", "racist"],
    "Service attentiveness / efficiency": ["attentive", "efficiency", "crew", "helpful", "staff"],
    "Tickets and refunds": ["ticketing","ticket", "refund", "cancellation", "booking", "change", "redemption"],
    "Baggage delivery": ["baggage", "luggage", "lost", "bag"],
    "Entertainment": ["entertainment", "movies", "movie", "film", "tv", "games"],
    "Online booking and check-in services": ["online", "check-in", "web"],
    "Boarding assistance": ["boarding", "assist", "delayed", "delay", "late"],
    "Airline Lounge : product facilities": ["lounge", "facilities", "airport"],
    "Cabin comfort & amenities" : ["cold", "blanket", "toothbrush", "towel", "pillow","earphone" ],
    "Scoot Airline" : ["scoot"],
    "General Topic": []  # Default topic
}

# Step 2: Map topic_label with target_topics
def map_topic_label(topic_label, target_topics):
    for topic, keywords in target_topics.items():
        if any(keyword in topic_label.lower() for keyword in keywords):
            return topic
    return "General Topic"

# Apply the mapping to the filtered BERTopic DataFrame
bertopic_mapped_df['mapped_topic'] = bertopic_mapped_df['topic_label'].apply(
    lambda x: map_topic_label(str(x), target_topics)
)

# Step 3: Add mapped_topic column to the DataFrame
print(f"BERTopic mapped topics processed: {len(bertopic_mapped_df)} reviews")

# Step 4: Focus on topic-1 in BERTopic and merge with Zero-Shot results
topic_minus_one_df = bertopic_df[bertopic_df['topic'] == -1].copy()

# Merge predicted_topic and confidence from Zero-Shot using matching criteria
merged_topic_minus_one = pd.merge(
    topic_minus_one_df,
    zero_shot_df[['published_date', 'cleaned_review', 'predicted_topic', 'confidence']],
    on=['published_date', 'cleaned_review'],  # Adjust columns if necessary
    how='left'
)

# Step 5: Duplicate predicted_topic into mapped_topic for topic-1 reviews
merged_topic_minus_one['mapped_topic'] = merged_topic_minus_one['predicted_topic']

print(f"Mapped Zero-Shot topics for topic -1: {len(merged_topic_minus_one)} reviews")

# Combine both mapped datasets
final_combined_df = pd.concat([bertopic_mapped_df, merged_topic_minus_one], ignore_index=True)

# Save the final results
final_combined_df.to_csv('final_mapped_topics_results.csv', index=False)
print("Final mapped topics saved successfully as 'final_mapped_topics_results.csv'")
```

```
BERTopic mapped topics processed: 4304 reviews
Mapped Zero-Shot topics for topic -1: 5696 reviews
Final mapped topics saved successfully as 'final_mapped_topics_results.csv'
```

Week 8 Topic Forcasting with LSTM and Prophet Models

1. Model Training

1.1 Training with LSTM (Long Short-Term Memory)

```
import pandas as pd

# Load the dataset
file_path = 'final_mapped_topics_results.csv'
df = pd.read_csv(file_path)

# Convert 'published_date' to datetime and handle timezone information
df['published_date'] = pd.to_datetime(df['published_date'], utc=True, errors='coerce')

# Group data by Date, Topic, and Sentiment
daily_topic_sentiment_counts = df.groupby(
    [df['published_date'].dt.date, 'mapped_topic', 'sentiment']
).size().reset_index(name='Count')

# Calculate total counts of all topics for each day and sentiment
daily_total_counts = daily_topic_sentiment_counts.groupby(
    ['published_date', 'sentiment']
)['Count'].sum().reset_index(name='Total_Count')

# Merge total counts back to the topic-level data
daily_topic_sentiment_counts = pd.merge(
    daily_topic_sentiment_counts,
    daily_total_counts,
    on=['published_date', 'sentiment']
)

# Calculate percentage for each topic per day per sentiment
daily_topic_sentiment_counts['Percentage'] = (
    daily_topic_sentiment_counts['Count'] / daily_topic_sentiment_counts['Total_Count'] * 100
)

# Sort the data for better readability
daily_topic_sentiment_counts_sorted = daily_topic_sentiment_counts.sort_values(
    by=['published_date', 'mapped_topic', 'sentiment']
)

# Save the results to a new CSV file for further analysis
output_file_path = 'daily_topic_sentiment_percentages.csv'  # Replace with your desired output path
daily_topic_sentiment_counts_sorted.to_csv(output_file_path, index=False)

# Display the resulting DataFrame for verification
print(daily_topic_sentiment_counts_sorted.head())
```

```
  published_date             mapped_topic sentiment  Count  Total_Count  \
```

```python
import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense
import matplotlib.pyplot as plt

# Load the preprocessed data
file_path = 'daily_topic_sentiment_percentages.csv'  # Replace with your file path
df = pd.read_csv(file_path)

# Function to prepare data and train LSTM for each topic
def train_lstm_for_topic(topic, sentiment, sequence_length=30):
    # Filter data for a specific topic and sentiment
    filtered_df = df[(df['mapped_topic'] == topic) & (df['sentiment'] == sentiment)]

    # Check if there is enough data for training
    if len(filtered_df) <= sequence_length:
        print(f"Not enough data for Topic: {topic}, Sentiment: {sentiment}")
        return None, None, None

    # Ensure data is sorted by date
    filtered_df['published_date'] = pd.to_datetime(filtered_df['published_date'])
    filtered_df = filtered_df.sort_values('published_date')

    # Extract percentage values
    percentage_values = filtered_df['Percentage'].values.reshape(-1, 1)

    # Normalize data
    scaler = MinMaxScaler()
    percentage_scaled = scaler.fit_transform(percentage_values)

    # Create sequences
    X, y = [], []
    for i in range(len(percentage_scaled) - sequence_length):
        X.append(percentage_scaled[i:i+sequence_length])
        y.append(percentage_scaled[i+sequence_length])
    X, y = np.array(X), np.array(y)

    # Build the LSTM model
    model = Sequential([
        LSTM(50, return_sequences=True, input_shape=(X.shape[1], X.shape[2])),
        LSTM(50),
        Dense(1)
    ])
    model.compile(optimizer='adam', loss='mse')

    # Train the model
    model.fit(X, y, epochs=20, batch_size=32, verbose=0)

    # Predict using the model
    predictions = model.predict(X)

    # Reverse normalization for predictions and actual values
    predicted_values = scaler.inverse_transform(predictions)
    actual_values = scaler.inverse_transform(y.reshape(-1, 1))

    return filtered_df.iloc[sequence_length:], actual_values, predicted_values

# Define top 3 topics for each sentiment
top_positive_topics = [
    'Service attentiveness / efficiency',
    'Meal service efficiency',
    'Seat comfort'
]
top_negative_topics = [
    'Service attentiveness / efficiency',
    'Tickets and refunds',
    'Seat comfort'
]

# Function to plot results for multiple topics
def plot_results(topics, sentiment, colors):
    plt.figure(figsize=(14, 7))
    for idx, topic in enumerate(topics):
        # Train and forecast for each topic
        dates, actual_values, predicted_values = train_lstm_for_topic(topic, sentiment)

        if dates is None:  # Skip topics with insufficient data
            continue

        # Plot actual values
        plt.plot(dates['published_date'], actual_values, linestyle='--', label=f'{topic} - Actual',
                 color=colors[idx], alpha=0.3, linewidth=1)

        # Plot predicted values
        plt.plot(dates['published_date'], predicted_values, label=f'{topic} - Predicted',
                 color=colors[idx], alpha=1, linewidth=2)

    # Add labels and legend
    plt.title(f'LSTM Forecasting for Top 3 Topics ({sentiment.capitalize()})')
    plt.xlabel('Date')
    plt.ylabel('Percentage')
    plt.legend(loc='best', title='Topics')
    plt.grid()
    plt.show()

# Define colors for each topic
colors = ['blue', 'orange', 'green']

# Plot for positive sentiment
plot_results(top_positive_topics, 'positive', colors)

# Plot for negative sentiment
plot_results(top_negative_topics, 'negative', colors)
```

1.2 Training with Prophet

3]:

```python
import numpy as np
import pandas as pd
from prophet import Prophet
import matplotlib.pyplot as plt
from sklearn.metrics import mean_squared_error, mean_absolute_error

# Load the preprocessed data
file_path = 'daily_topic_sentiment_percentages.csv'  # Replace with your file path
df = pd.read_csv(file_path)

# Function to prepare data and train Prophet for each topic
def train_prophet_for_topic(topic, sentiment):
    # Filter data for a specific topic and sentiment
    filtered_df = df[(df['mapped_topic'] == topic) & (df['sentiment'] == sentiment)]

    # Check if there is enough data for training
    if len(filtered_df) <= 10:  # Ensure sufficient data
        print(f"Not enough data for Topic: {topic}, Sentiment: {sentiment}")
        return None, None, None

    # Prepare data for Prophet
    filtered_df['published_date'] = pd.to_datetime(filtered_df['published_date'])
    filtered_df = filtered_df.sort_values('published_date')
    prophet_df = filtered_df[['published_date', 'Percentage']].rename(
        columns={'published_date': 'ds', 'Percentage': 'y'}
    )

    # Train the Prophet model
    model = Prophet()
    model.fit(prophet_df)

    # Create a dataframe for future predictions
    future = model.make_future_dataframe(periods=30)  # Forecast next 30 days
    forecast = model.predict(future)

    # Match predictions with actual values
    actual_values = prophet_df['y']
    predicted_values = forecast.loc[:len(actual_values)-1, 'yhat']  # Match with historical data

    return prophet_df, actual_values, predicted_values

# Define top 3 topics for each sentiment
top_positive_topics = [
    'Service attentiveness / efficiency',
    'Meal service efficiency',
    'Seat comfort'
]
top_negative_topics = [
    'Service attentiveness / efficiency',
    'Tickets and refunds',
    'Seat comfort'
]

# Function to plot results for multiple topics
def plot_results(topics, sentiment, colors):
    plt.figure(figsize=(14, 7))
    for idx, topic in enumerate(topics):
        # Train and forecast for each topic
        data, actual_values, predicted_values = train_prophet_for_topic(topic, sentiment)

        if data is None:  # Skip topics with insufficient data
            continue

        # Plot actual values
        plt.plot(data['ds'], actual_values, linestyle='--', label=f'{topic} - Actual',
                 color=colors[idx], alpha=0.3, linewidth=1)

        # Plot predicted values
        plt.plot(data['ds'], predicted_values, label=f'{topic} - Predicted',
                 color=colors[idx], alpha=1, linewidth=2)

    # Add labels and legend
    plt.title(f'Prophet Forecasting for Top 3 Topics ({sentiment.capitalize()})')
    plt.xlabel('Date')
    plt.ylabel('Percentage')
    plt.legend(loc='best', title='Topics')
    plt.grid()
    plt.show()

# Define colors for each topic
colors = ['blue', 'orange', 'green']

# Plot for positive sentiment
plot_results(top_positive_topics, 'positive', colors)

# Plot for negative sentiment
plot_results(top_negative_topics, 'negative', colors)
```

2. Model Evaluation

```
]: from sklearn.metrics import mean_squared_error, mean_absolute_error
import numpy as np
import pandas as pd

# Function to calculate RMSE, MAE, and MAPE for each topic
def evaluate_metrics(topics, sentiment, method, train_model_function):
    results = []

    for idx, topic in enumerate(topics):
        print(f"Evaluating {method} on topic: {topic}")

        # Train and forecast using the provided model function
        dates, actual_values, predicted_values = train_model_function(topic, sentiment)

        if dates is None:  # Skip topics with insufficient data
            print(f"Not enough data for topic: {topic}")
            continue

        # Evaluate performance
        rmse = np.sqrt(mean_squared_error(actual_values, predicted_values))
        mae = mean_absolute_error(actual_values, predicted_values)
        mape = np.mean(np.abs((actual_values - predicted_values) / actual_values)) * 100  # MAPE formula

        print(f"{method} - Topic: {topic}, RMSE: {rmse:.2f}, MAE: {mae:.2f}, MAPE: {mape:.2f}%")

        # Store results
        results.append({'Topic': topic, 'Method': method, 'RMSE': rmse, 'MAE': mae, 'MAPE': mape})

    return results

# Define topics and sentiments
top_positive_topics = [
    'Service attentiveness / efficiency',
    'Meal service efficiency',
    'Seat comfort'
]
top_negative_topics = [
    'Service attentiveness / efficiency',
    'Tickets and refunds',
    'Seat comfort'
]

# Evaluate Prophet
prophet_results = evaluate_metrics(
    topics=top_positive_topics,
    sentiment='positive',
    method='Prophet',
    train_model_function=train_prophet_for_topic  # Replace with Prophet training function
)

# Evaluate LSTM
lstm_results = evaluate_metrics(
    topics=top_positive_topics,
    sentiment='positive',
    method='LSTM',
    train_model_function=train_lstm_for_topic  # Replace with LSTM training function
)

# Combine results into a single DataFrame for comparison
all_results = pd.DataFrame(prophet_results + lstm_results)

# Display results
print(all_results)

# Optionally, save to CSV for reporting
all_results.to_csv("metrics_comparison_results_with_mape.csv", index=False)
```

3. Top 3 Topics Forcasting

3.1 Recall LSTM Result

```
import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense
import matplotlib.pyplot as plt

# Load the preprocessed data
file_path = 'daily_topic_sentiment_percentages.csv'
df = pd.read_csv(file_path)

# Function to prepare data and train LSTM for each topic
def train_lstm_for_topic(topic, sentiment, sequence_length=30):
    # Filter data for a specific topic and sentiment
    filtered_df = df[(df['mapped_topic'] == topic) & (df['sentiment'] == sentiment)]

    # Check if there is enough data for training
    if len(filtered_df) <= sequence_length:
        print(f"Not enough data for Topic: {topic}, Sentiment: {sentiment}")
        return None, None, None

    # Ensure data is sorted by date
    filtered_df['published_date'] = pd.to_datetime(filtered_df['published_date'])
    filtered_df = filtered_df.sort_values('published_date')

    # Extract percentage values
    percentage_values = filtered_df['Percentage'].values.reshape(-1, 1)

    # Normalize data
    scaler = MinMaxScaler()
    percentage_scaled = scaler.fit_transform(percentage_values)

    # Create sequences
    X, y = [], []
    for i in range(len(percentage_scaled) - sequence_length):
        X.append(percentage_scaled[i:i+sequence_length])
        y.append(percentage_scaled[i+sequence_length])
    X, y = np.array(X), np.array(y)

    # Build the LSTM model
    model = Sequential([
        LSTM(50, return_sequences=True, input_shape=(X.shape[1], X.shape[2])),
        LSTM(50),
        Dense(1)
    ])
    model.compile(optimizer='adam', loss='mse')

    # Train the model
    model.fit(X, y, epochs=20, batch_size=32, verbose=0)  # Suppress detailed output

    # Predict using the model
    predictions = model.predict(X)

    # Reverse normalization for predictions and actual values
    predicted_values = scaler.inverse_transform(predictions)
    actual_values = scaler.inverse_transform(y.reshape(-1, 1))

    return filtered_df.iloc[sequence_length:], actual_values, predicted_values

# Define top 3 topics for each sentiment
top_positive_topics = [
    'Service attentiveness / efficiency',
    'Meal service efficiency',
    'Seat comfort'
]
top_negative_topics = [
    'Service attentiveness / efficiency',
    'Tickets and refunds',
    'Seat comfort'
]

# Function to plot results for multiple topics
def plot_results(topics, sentiment, colors):
    plt.figure(figsize=(14, 7))
    for idx, topic in enumerate(topics):
        # Train and forecast for each topic
        dates, actual_values, predicted_values = train_lstm_for_topic(topic, sentiment)

        if dates is None:  # Skip topics with insufficient data
            continue

        # Plot actual values
        plt.plot(dates['published_date'], actual_values, linestyle='--', label=f'{topic} - Actual',
                 color=colors[idx], alpha=0.3, linewidth=1)

        # Plot predicted values
        plt.plot(dates['published_date'], predicted_values, label=f'{topic} - Predicted',
                 color=colors[idx], alpha=1, linewidth=2)

    # Add labels and legend
    plt.title(f'LSTM Forecasting for Top 3 Topics ({sentiment.capitalize()})')
    plt.xlabel('Date')
    plt.ylabel('Percentage')
    plt.legend(loc='best', title='Topics')
    plt.grid()
    plt.show()

# Define colors for each topic
colors = ['blue', 'orange', 'green']

# Plot for positive sentiment
plot_results(top_positive_topics, 'positive', colors)

# Plot for negative sentiment
plot_results(top_negative_topics, 'negative', colors)
```

3.2 Topic Forcasting

```
]: import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense
import matplotlib.pyplot as plt
from datetime import timedelta

# Load the preprocessed data
file_path = 'daily_topic_sentiment_percentages.csv'
df = pd.read_csv(file_path)

# Function to prepare data and train LSTM for each topic
def train_lstm_for_topic(topic, sentiment, sequence_length=30):
    # Filter data for a specific topic and sentiment
    filtered_df = df[(df['mapped_topic'] == topic) & (df['sentiment'] == sentiment)]

    # Check if there is enough data for training
    if len(filtered_df) <= sequence_length:
        print(f"Not enough data for Topic: {topic}, Sentiment: {sentiment}")
        return None, None, None

    # Ensure data is sorted by date
    filtered_df['published_date'] = pd.to_datetime(filtered_df['published_date'])
    filtered_df = filtered_df.sort_values('published_date')

    # Extract percentage values
    percentage_values = filtered_df['Percentage'].values.reshape(-1, 1)

    # Normalize data
    scaler = MinMaxScaler()
    percentage_scaled = scaler.fit_transform(percentage_values)

    # Create sequences
    X, y = [], []
    for i in range(len(percentage_scaled) - sequence_length):
        X.append(percentage_scaled[i:i+sequence_length])
        y.append(percentage_scaled[i+sequence_length])
    X, y = np.array(X), np.array(y)

    # Build the LSTM model
    model = Sequential([
        LSTM(50, return_sequences=True, input_shape=(X.shape[1], X.shape[2])),
        LSTM(50),
        Dense(1)
    ])
    model.compile(optimizer='adam', loss='mse')

    # Train the model
    model.fit(X, y, epochs=20, batch_size=32, verbose=0)

    # Predict using the model
    predictions = model.predict(X)

    # Reverse normalization for predictions and actual values
    predicted_values = scaler.inverse_transform(predictions)
    actual_values = scaler.inverse_transform(y.reshape(-1, 1))

    return filtered_df.iloc[sequence_length:], actual_values, predicted_values, model, scaler

# Function to forecast future percentages using the trained model
def forecast_future(model, scaler, last_sequence, start_date, forecast_horizon):
    # Prepare future dates
    future_dates = [start_date + timedelta(days=i) for i in range(1, forecast_horizon + 1)]
```

```python
    # Forecast future values
    predictions = []
    for _ in range(forecast_horizon):
        # Reshape the last sequence for prediction
        last_sequence_reshaped = last_sequence.reshape(1, last_sequence.shape[0], 1)
        pred = model.predict(last_sequence_reshaped, verbose=0)
        predictions.append(pred[0, 0])
        # Update the last sequence with the new prediction
        last_sequence = np.append(last_sequence[1:], pred, axis=0)

    # Reverse normalization for predictions
    predictions = scaler.inverse_transform(np.array(predictions).reshape(-1, 1)).flatten()

    return future_dates, predictions

# Function to extend forecast and plot results
def extend_forecast_plot(topics, sentiment, colors):
    plt.figure(figsize=(14, 7))
    for idx, topic in enumerate(topics):
        # Train and get the last sequence, model, and scaler
        dates, actual_values, predicted_values, model, scaler = train_lstm_for_topic(topic, sentiment)

        if dates is None:  # Skip topics with insufficient data
            continue

        # Normalize the last sequence
        percentage_values = dates['Percentage'].values.reshape(-1, 1)
        percentage_scaled = scaler.transform(percentage_values)

        # Use the last 30 days as the last sequence
        if len(percentage_scaled) >= 30:
            last_sequence = percentage_scaled[-30:]
        else:
            print(f"Not enough data to create a sequence for topic: {topic}, sentiment: {sentiment}")
            continue

        # Set start date and forecast horizon
        start_date = dates['published_date'].iloc[-1]
        forecast_horizon = (pd.Timestamp('2025-12-31') - start_date).days

        # Forecast future values
        future_dates, future_predictions = forecast_future(model, scaler, last_sequence, start_date, forecast_horizo

        # Plot historical actual and predicted values
        plt.plot(dates['published_date'], actual_values, linestyle='--', label=f'{topic} - Actual', color=colors[idx
        plt.plot(dates['published_date'], predicted_values, label=f'{topic} - Predicted', color=colors[idx], alpha=1

        # Plot future predictions
        plt.plot(future_dates, future_predictions, label=f'{topic} - Forecast', color=colors[idx], linewidth=2)

    # Customize the plot
    plt.title(f'LSTM Extended Forecasting for Top 3 Topics ({sentiment.capitalize()}) until Dec 2025')
    plt.xlabel('Date')
    plt.ylabel('Percentage')
    plt.legend(loc='best', title='Topics')
    plt.grid()
    plt.show()

# Define top 3 topics for each sentiment
top_positive_topics = [
    'Service attentiveness / efficiency',
    'Meal service efficiency',
    'Seat comfort'
]
top_negative_topics = [
    'Service attentiveness / efficiency',
    'Tickets and refunds',
    'Seat comfort'
]

# Define colors for each topic
colors = ['blue', 'orange', 'green']

# Extend forecast for positive sentiment
print("Extending forecast for Positive Sentiments...")
extend_forecast_plot(top_positive_topics, 'positive', colors)

# Extend forecast for negative sentiment
print("Extending forecast for Negative Sentiments...")
extend_forecast_plot(top_negative_topics, 'negative', colors)
```