Exploring Customer Feedback: Sentiment and Service Quality in Indian Airlines

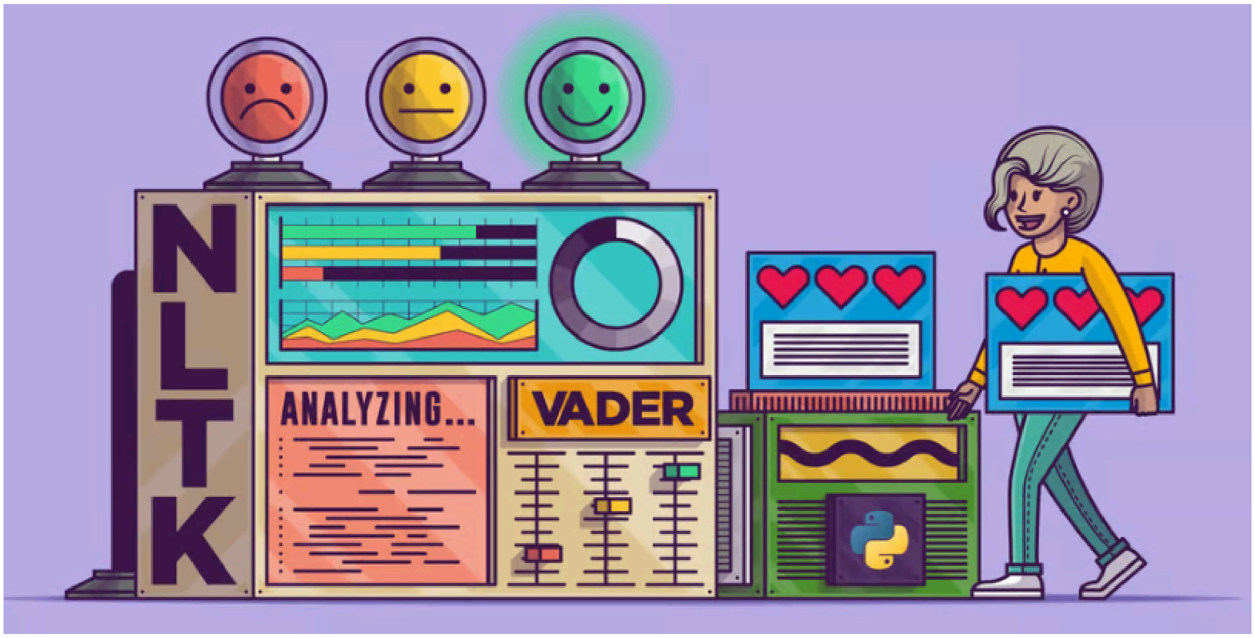


Table of Contents

List of Tables

List of Diagrams

List of Figures

1. Introduction

Sentiment analysis, commonly known as opinion mining, is defined as a task to detect, extract and classify people's opinions on something. It is an essential tool for understanding consumers' experiences and improving them. By analyzing the feedback received from different sources, companies can potentially gain insights into their customers preferences, perceptions, emotions and pain points (Banerji et al., 2023).

The airline industry is a highly competitive business, and can greatly benefit from learning about their customers' experiences. The insights will enable them to improve their services and increase the customer stickiness in the long run. This dynamic landscape requires a deep understanding of the factors that influence passenger choices. Key elements such as service quality, pricing strategies, and overall customer satisfaction significantly shape consumer preferences (Banerji et al., 2023; Ozkul et al., 2020).

This study focuses on four major airlines in India—Air India, Indigo, SpiceJet, and Vistara—analyzing over 3,700 reviews sourced from the reputable Skytrax website. Known for its unbiased passenger feedback, Skytrax serves as an ideal platform for this research. The selection of these airlines is driven by several considerations. First, they represent the Indian aviation market, including both full-service and low-cost carriers. This diversity enables a comprehensive analysis of varying customer experiences. Second, these airlines hold significant market shares and have different reputations, providing a rich dataset for sentiment analysis.

Reviews from the website are scraped using Beautiful Soup then cleaning and organizing them in Python. Text cleaning is done by removing unnecessary words and characters with NLTK. The cleaned text is then analyzed using VADER, which categorizes reviews into positive, neutral, or negative sentiment. Visualizations are created to clearly present the results.

To further investigate the relationships among identified variables related to airline service quality and customer satisfaction, this report applied Partial Least Squares Structural Equation Modeling (PLS-SEM) which studies both the measurement and structural model. This statistical technique is well-suited for complex models involving multiple latent variables (Hair et al., 2013). In this context, we will focus on three primary latent variables: Service Quality, which encompasses dimensions like seating comfort, food quality, and entertainment; Customer Satisfaction, which includes value for money and sentiment; and Operational Efficiency, which considers ground service and staff performance. The overall rating will serve as a comprehensive measure of passenger satisfaction. We aim to examine the cause effect relationship between the three latent variables and overall rating. This will help airlines understand what is driving customer preferences and behaviors, and can form strategic decisions to improve customer loyalty, brand image, and growth.

Objectives

The objectives of this study are:

- To extract and examine the Airline reviews and ratings derived from the website, Skytrax for the decade 2014- 2024.
- To conduct sentiment analysis on the reviews.
- To explore the impact of airline attributes, including service quality and operational efficiency, on customer satisfaction and overall airline ratings.

2. Methodology

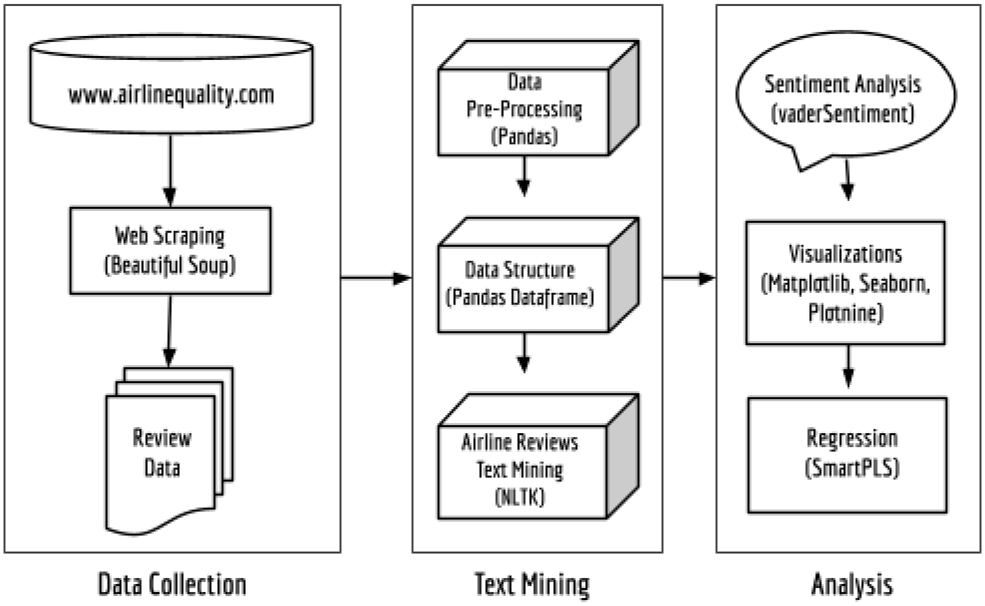


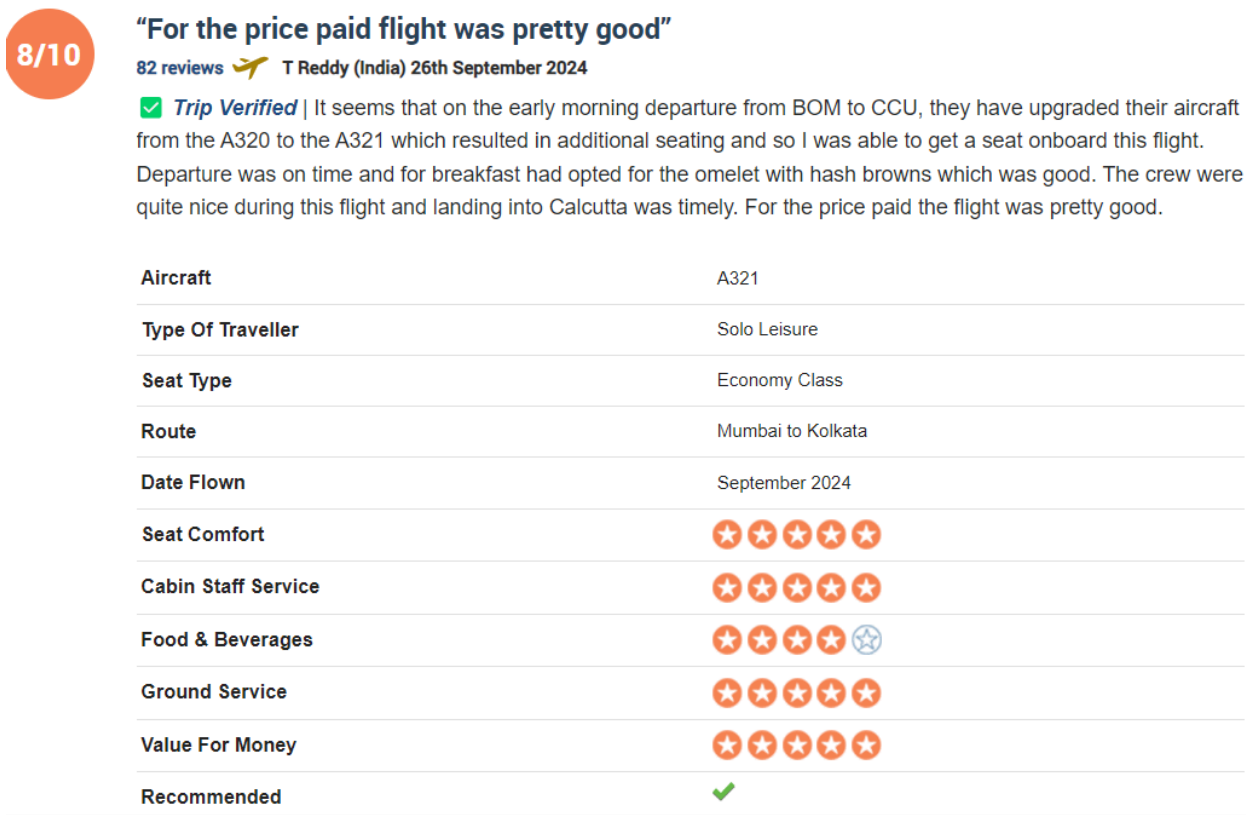
Diagram 1: Methodology Flowchart

2.1 Data Collection

The customer reviews were extracted from the Skytrax Airline Quality Reviews website, which is said to feature genuine reviews from passengers across multiple airlines all over the world. The terms and conditions of Skytrax allows users to freely use their data unless it is for commercial purposes. We employed the Beautiful Soup library in Python to scrape customer reviews data from this website. Beautiful Soup allowed us to extract the review text, rating scores, and other relevant fields. We focused on reviews of four major airlines operating in India, namely Air India (726 reviews), Indigo (829 reviews), Spice Jet (1020 reviews), and Vistara (1174 reviews) over the last decade (2014 - 2024). In total, the data scraped amounted to 3745 records or reviews after cleaning.

Review Format of Skytrax

The following figure is an example review from the Skytrax website. It consists of both, text review and ratings of different parameters. There is a review header which is the title of the review, an overall rating, date of review, details of the travel like demographic information and five different airline attribute ratings.



8/10

"For the price paid flight was pretty good"

82 reviews T Reddy (India) 26th September 2024

Trip Verified | It seems that on the early morning departure from BOM to CCU, they have upgraded their aircraft from the A320 to the A321 which resulted in additional seating and so I was able to get a seat onboard this flight. Departure was on time and for breakfast had opted for the omelet with hash browns which was good. The crew were quite nice during this flight and landing into Calcutta was timely. For the price paid the flight was pretty good.

Aircraft	A321
Type Of Traveller	Solo Leisure
Seat Type	Economy Class
Route	Mumbai to Kolkata
Date Flown	September 2024
Seat Comfort	★★★★★
Cabin Staff Service	★★★★★
Food & Beverages	★★★★☆
Ground Service	★★★★★
Value For Money	★★★★★
Recommended	✓

Diagram 2: Example of Skytrax Review

Source: https://www.airlinequality.com/airline-reviews/indigo-airlines

2.2 Data Preprocessing

To clean and prepare the data obtained for analysis we used Python Pandas, and undertook the following steps:

Columns Removed: The columns that were not required for the analysis were dropped such as "Status" and "WiFi".

Handling Missing Data: All observations consisting null values were removed so as to maintain the quality and consistency of the dataset.

Combining Airline Datasets: The datasets for the four different airlines were appended, stacking all the reviews into a single dataset to enable for a broader comparative analysis.

Recommendation Column Recoded: The "Recommendation" column, originally with "Yes/No" values, was re-coded to binary values (1 for "Yes" and 0 for "No") for easier analysis.

Date Format Standardization: The date field was standardized, and month and year were extracted into separate columns.

2.3 Text Mining and Sentiment Analysis

Python has a lot of available libraries that assist in efficient text analytics. After using Beautiful Soup library for extracting all the necessary texts and numbers, and processing it using Pandas, the collected data stored in a Panda's Dataframe was used for sentiment analysis. Text mining and Sentiment analysis is performed only on the review text using the NLTK library and vaderSentiment in Python respectively.

To prepare the review for sentiment analysis, we removed common stop words, punctuation and special characters. The cleaned text data was then processed using vaderSentiment library to generate compound scores and segregate the reviews into the following sentiments:

Positive Sentiment: Reviews that express satisfaction and appreciation.

Neutral Sentiment: Reviews with no strong positive or negative tone.

Negative Sentiment: Reviews that highlight problems or dissatisfaction.

The results of the analysis are then visualized in the result section and used for further cause- effect analysis in section 2.4.

2.4 Regression using PLS-SEM

Partial Least Squares Structural Equation Modeling (PLS-SEM) analysis was employed to analyze the collected data and the sentiment data. PLS-SEM is good for analyzing data from a secondary source with many items that may or may not follow normal distribution rule, and it further permits an unrestricted use of single item and formative measures (Hair, Risher, Sarstedt, & Ringle, 2019). The following figure shows the conceptual model for the study. In the model, service quality and operational efficiency are exogenous variables, customer satisfaction is a mediating variable, and overall rating is an endogenous variable.

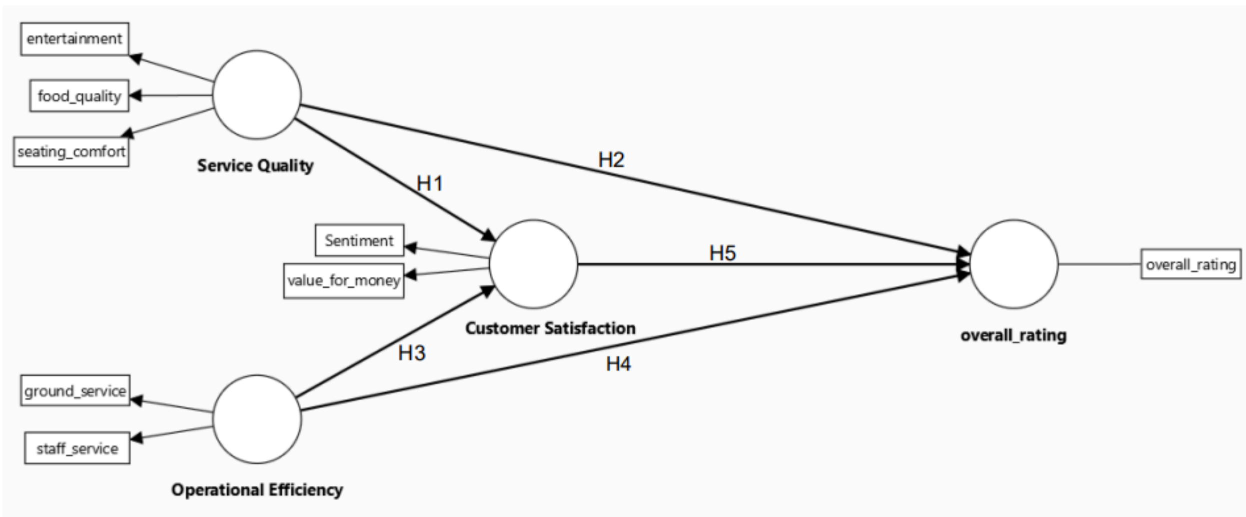


Diagram 3: Conceptual model

Source: Designed Using SmartPls

The measurement model is assessed on the basis of the reliability and validity of the constructs. The reflective measures are evaluated based on their composite reliability, convergent validity, and discriminant validity (Hair et al., 2012).

3. Result

This section presents the results of the sentiment analysis on airline reviews, which includes both numerical ratings and textual feedback. Various data visualization techniques were employed to understand the customer sentiments regarding the airline service, followed by text mining and sentiment analysis of the review text.

We began by loading the dataset, which contained detailed airline reviews with attributes such as travel class, route, date of travel, and individual ratings for factors like seating comfort, staff service, food quality, and entertainment. These variables were crucial in helping us identify patterns in customer satisfaction across different airlines and travel classes.

3.1 Visualization of Ratings and Service Aspect

To get a broad understanding of the overall customer experience, we first analyzed the average ratings of airlines. A bar plot of the average overall rating by airlines revealed significant differences in performance.

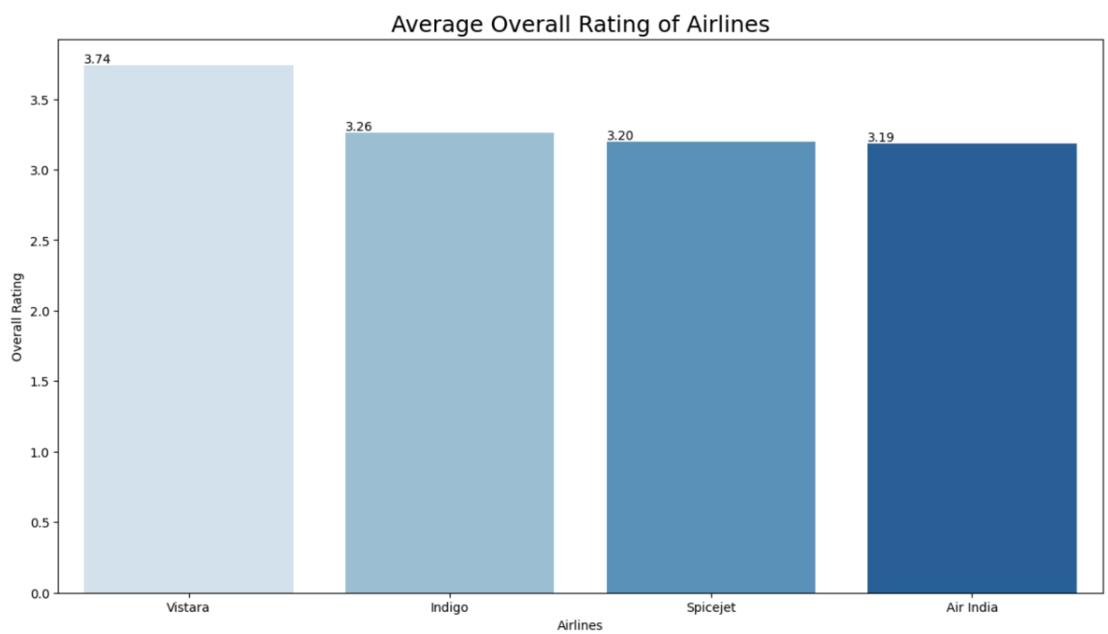


Figure 1: Average Overall Rating of Airlines

According to figure 1, Airlines like Vistara and Indigo stood out with the highest overall ratings, while others had lower averages. This suggests that certain airlines consistently provide a better service, as perceived by passengers. Vistara's strong performance can be attributed to its superior business class service, which we will explore further.

In figure 2, the analysis of seating comfort followed a similar pattern. Airlines that performed well in overall ratings, such as Vistara, also excelled in seating comfort. On the other hand, Spicejet struggled with lower seat comfort scores. Since seating comfort is a critical factor influencing a passenger's flight experience, airlines with low ratings in this area should focus on improving their seat design, especially for longer flights.

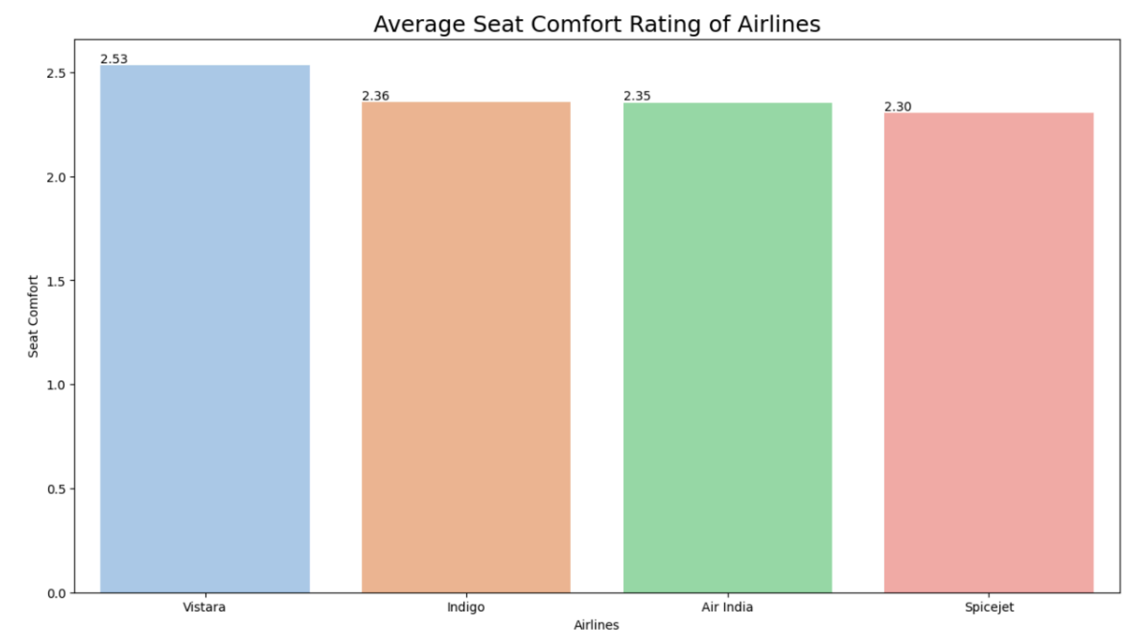


Figure 2: Average Seat Comfort Rating of Airlines

In figure 3, we then segmented the data by travel class to explore whether travel class influenced customer satisfaction. The results showed that Business Class and premium economy passengers consistently rated their experiences higher than Economy Class passengers, particularly for airlines like Vistara and SpiceJet. In contrast, Economy Class passengers often reported lower satisfaction, especially for airlines like Air India, where reviews revealed dissatisfaction with seat comfort, food quality, and staff service.

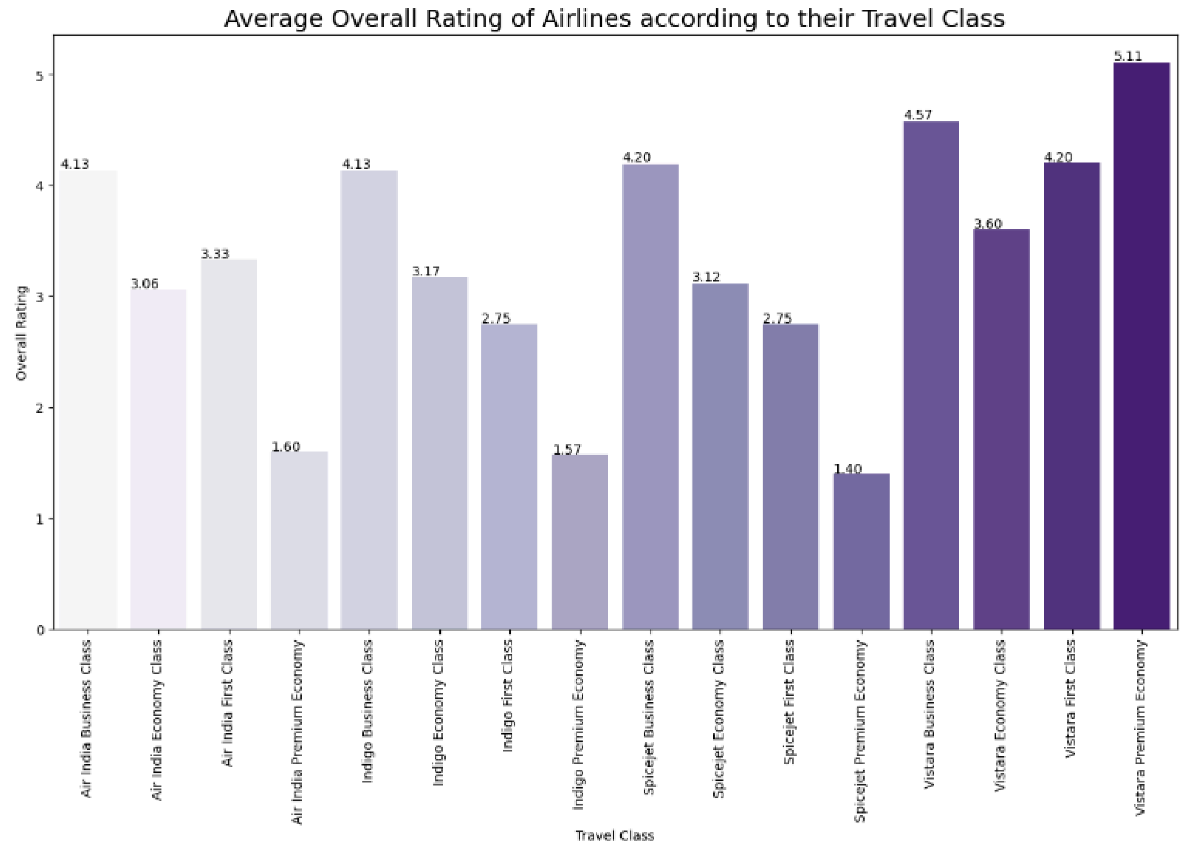


Figure 3: Average Overall Rating of Airlines According to Travel Class

These findings emphasize that premium classes offer a more satisfying experience, which may justify the higher ticket prices, while Economy Class presents an opportunity for airlines to make improvements.

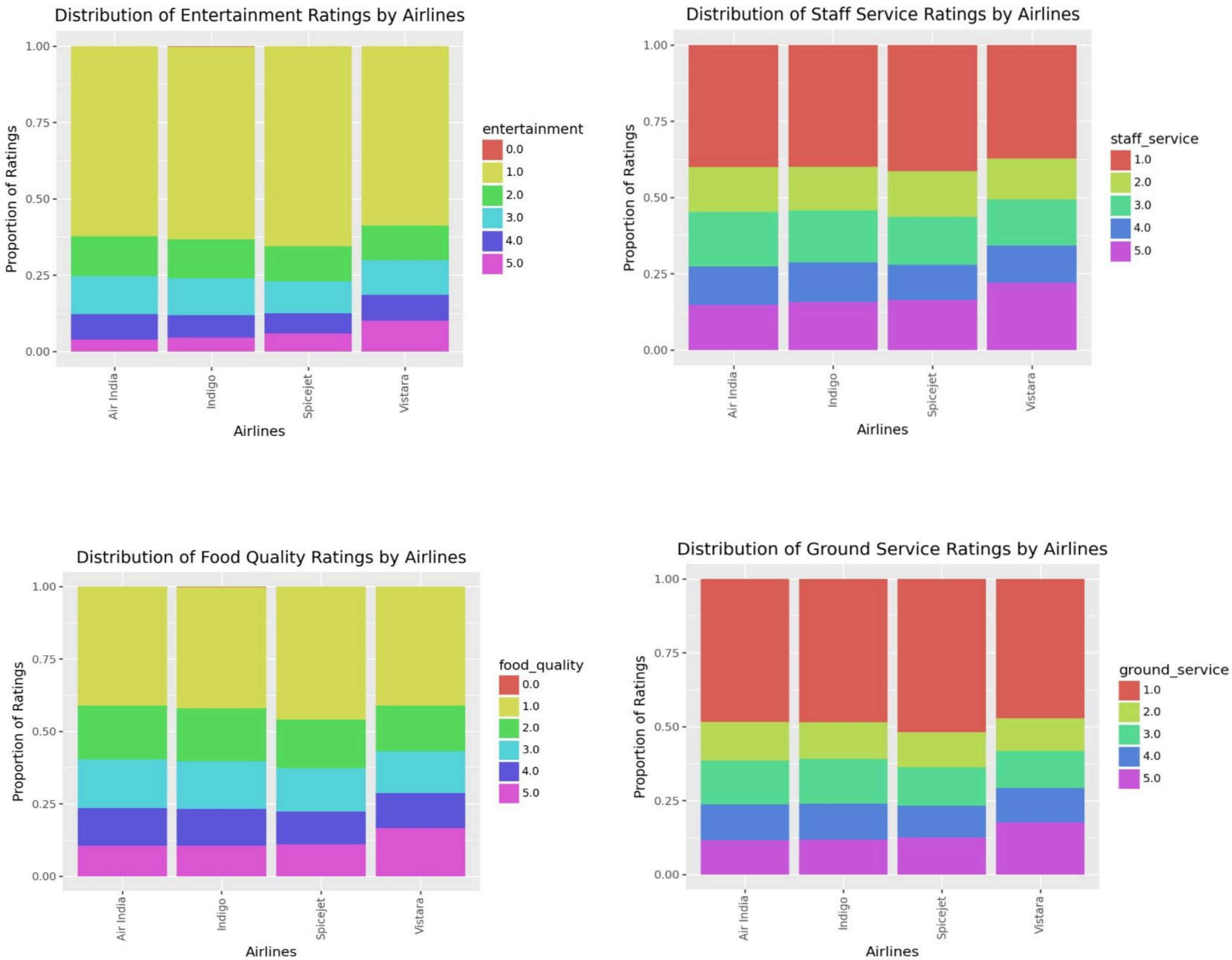


Figure 4: Distribution of Entertainment, Staff Service, Food Quality and Ground Service by Airlines

Figure 4 shows further breakdowns of ratings for ground service, staff service, and food quality showed that Vistara received positive ratings across the board, especially for staff service, while Air India and other budget carriers like Indigo scored poorly in food quality and staff behavior. The stacked bar plots visualizing these service aspects indicated that poor staff service and food quality are common issues for low-rated airlines, which could be key factors contributing to negative customer experiences.

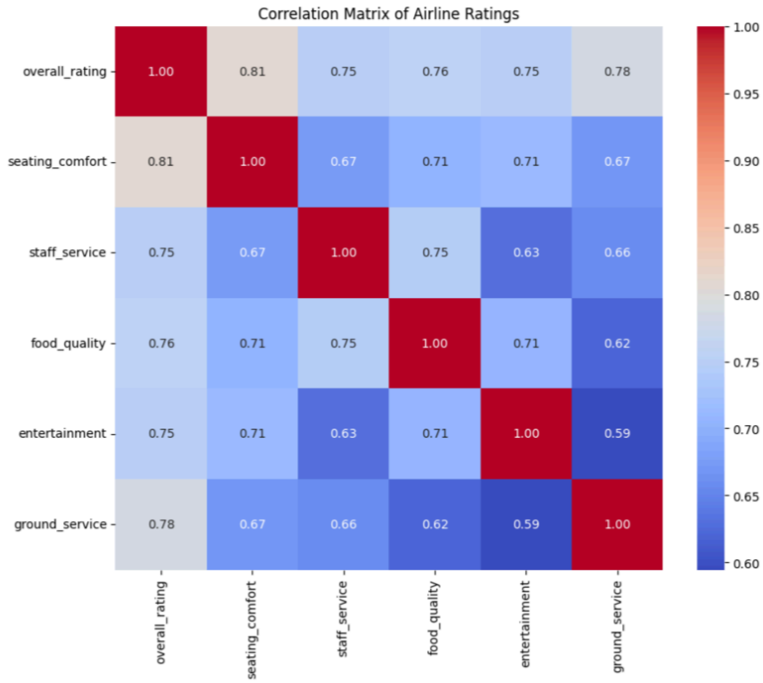


Figure 5: Correlation Matrix of Airline Ratings

To further substantiate these findings, a correlation matrix was plotted to assess the relationships between different service factors and overall ratings. According to figure 5, strong correlations were observed between overall ratings and factors like seating comfort, staff service, and food quality. This indicates that improving any of these critical areas could lead to a noticeable increase in overall customer satisfaction. Specifically, airlines that perform poorly in one of these aspects could benefit from targeted improvements to boost their reputation.

Figure 6 is a box plot that visualizes the overall customer ratings based on different travel classes: Economy Class, Business Class, Premium Economy, and First Class. From this plot, it's clear that Economy Class has the lowest overall rating, with its median rating notably lower than the other classes. Business Class, Premium Economy, and First Class all have similar ratings, though Business Class slightly edges out the others with a higher median.

Figure 7 is a violin plot that explores how different airlines—Indigo, Vistara, Air India, and SpiceJet—perform in terms of overall customer ratings. This plot not only shows the spread of the ratings but also visualizes the density of the data at different points of the rating scale.

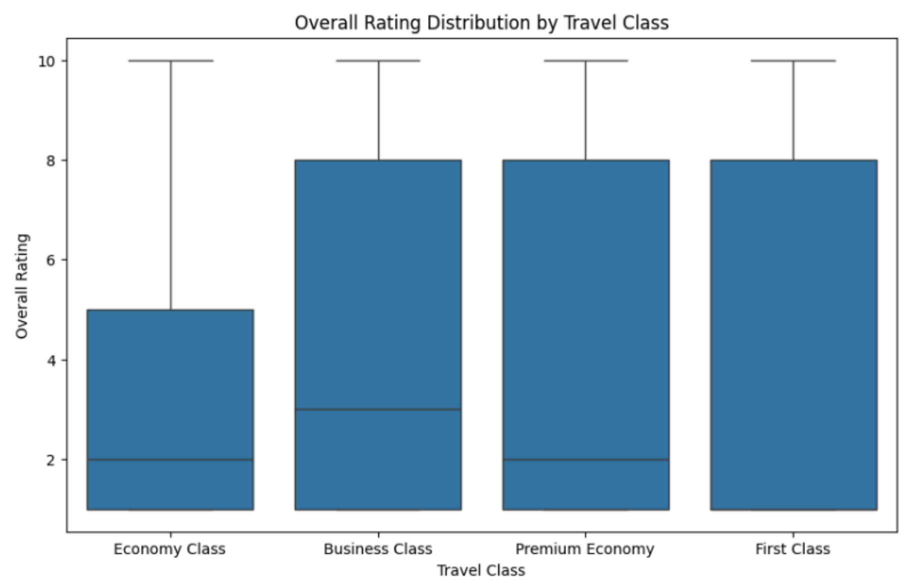


Figure 6: Overall Rating Distribution by Travel Class

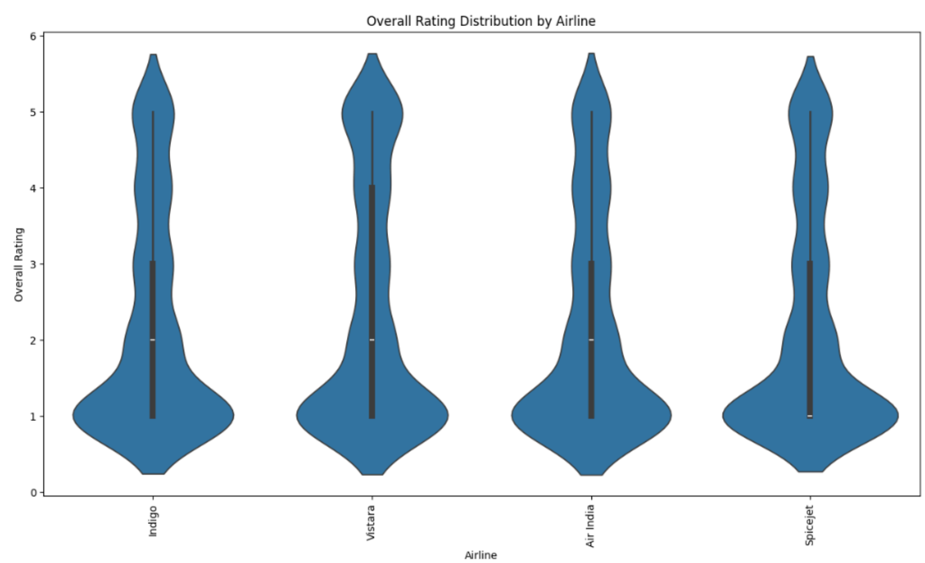


Figure 7: Overall Rating Distribution by Airline

Each airline's distribution follows a somewhat similar pattern, with the majority of ratings clustered around the middle (3 to 4 on the scale). Air India appears to have a wider spread of ratings compared to the others, indicating more variability in customer experiences.

Indigo and Vistara display more concentrated distributions, implying more consistent ratings from their customers.

3.2 Text Mining and Sentiment Analysis Using NLP

Following the visual analysis of ratings, we moved on to analyze the sentiment expressed in customer reviews using text mining techniques. The text data was cleaned by removing special characters, stopwords, and applying lemmatization, which normalized the words for better analysis. This allowed us to focus on the core content of the reviews without being influenced by irrelevant terms. The cleaned reviews were then analyzed using the VaderSentiment analyzer to classify each review as positive, negative, or neutral based on the compound sentiment score.

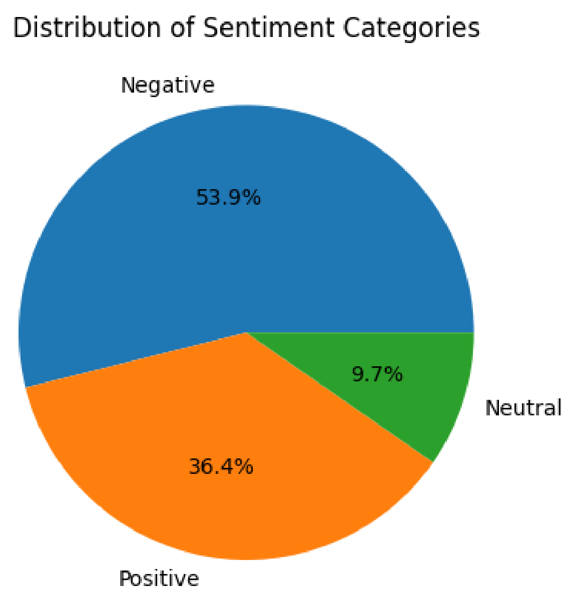


Figure 8: Distribution of Sentiment Categories (Pie Chart)

The results revealed a skew towards negative feedback, with 53.9% of the reviews being negative, while 36.5% were positive, and 9.6% were neutral. The pie chart and bar graph show a clear dominance of negative sentiments, suggesting widespread dissatisfaction among airline customers.

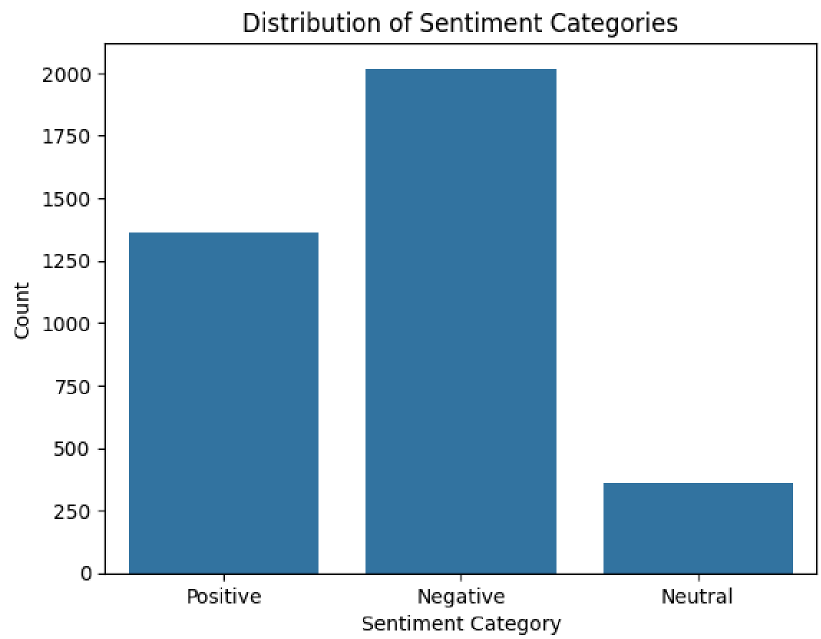


Figure 9: Distribution of Sentiment Categories (Bar Chart)

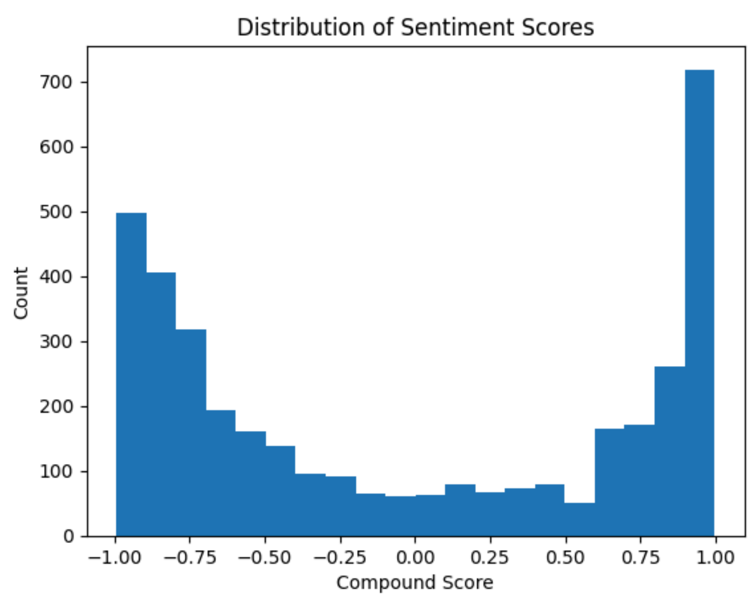


Figure 10: Distribution of Sentiment Scores

To better understand the intensity of customer opinions, a histogram of sentiment scores was plotted. It showed that most reviews were either strongly positive or strongly negative, with very few neutral sentiments. This suggests that passengers tend to leave reviews only when they have had an extreme experience, whether good or bad.

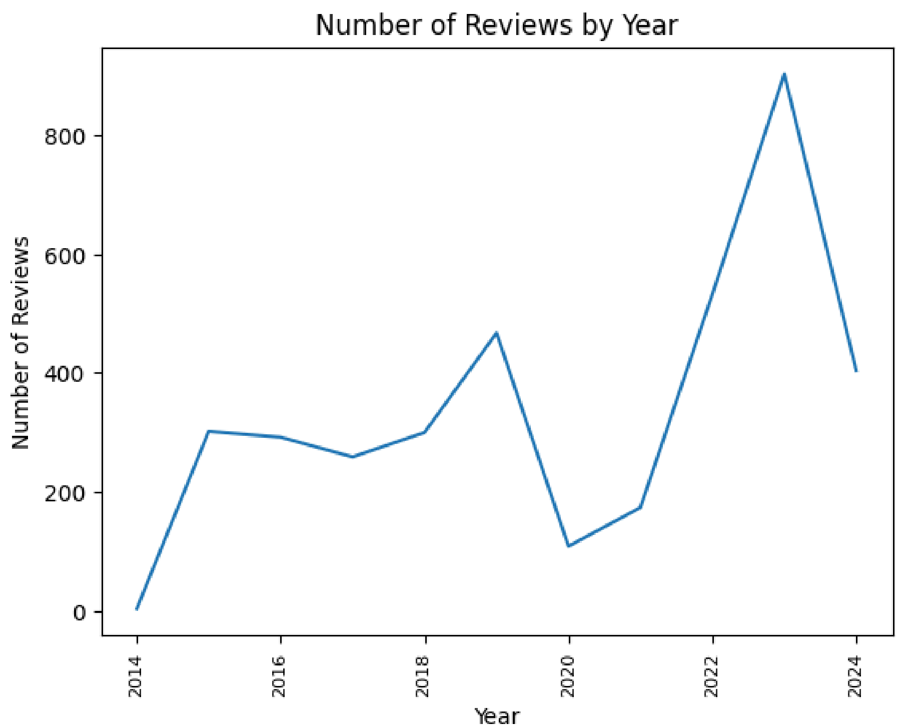


Figure 11: Number of Reviews by Year

This chart captures the trend in the frequency of airline reviews over time. From the chart, we can observe variations in the volume of reviews over the years, which might be influenced by several factors such as changes in airline operations, customer experiences, or even broader events like the COVID-19 pandemic that significantly impacted air travel. Peaks in certain years indicate periods of heightened customer feedback.

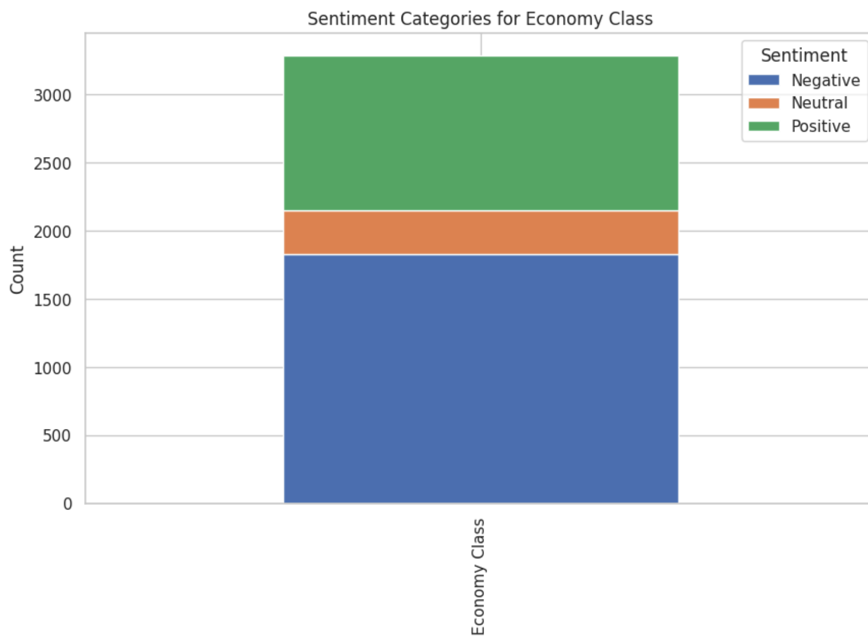


Figure 12: Sentiment Categories for Economy Class

Further, a closer examination of sentiment by travel class revealed that Economy Class passengers were significantly more likely to leave negative reviews, underscoring the need for service improvements in this class. The negative reviews commonly cited discomfort, poor food quality, and unresponsive staff as key issues.

Figure 13: Most Common Words in Reviews

A word cloud was created to visualize the most frequent words that arise in the reviews.

3.3 Model Results

Descriptive Statistics

Table 1: Descriptive Statistics

Name	Mean	Median	Scale min	Scale max	Standard deviation	Excess kurtosis	Skewness
seating_comfort	2.397	2	0	5	1.439	-1.093	0.554
staff_service	2.544	2	0	5	1.532	-1.317	0.429
food_quality	2.348	2	0	5	1.442	-1.019	0.624
entertainment	1.84	1	0	5	1.267	0.366	1.293
ground_service	2.291	2	0	5	1.498	-1.055	0.685
value_for_money	2.259	2	1	5	1.501	-0.957	0.763
recommended	0.26	0	0	1	0.438	-0.796	1.097
overall_rating	3.38	2	1	10	3.082	-0.422	1.051
Compound Score	-0.034	-0.21	-0.994	0.997	0.754	-1.684	0.157
Sentiment	-0.174	-1	-1	1	0.934	-1.762	0.353

Table no. 1 above shows the descriptive statistics of the variables in our study. The variable Sentiment is categorical ranging between -1 and 1, for positive (1), neutral (0), and negative (-1). The compound score is a metric generated using Python by analyzing the sentiments of the reviews. The mean compound score of the 3745 reviews is -0.034. The variables seating comfort, staff service, food quality, entertainment and ground service, are star ratings ranging from 0 to 5 stars.

Skewness measures the asymmetry of the distribution. A positive skewness indicates a longer tail to the right, while a negative skewness indicates a longer tail to the left. The "entertainment" variable has a positive skewness (1.293), suggesting that a few individuals gave very high ratings, pulling the mean to the right.

Standard Deviation measures the spread of the data around the mean. A higher standard deviation indicates greater variability. The "overall_rating" has the highest standard deviation (3.082), suggesting a wider range of responses. The overall rating, assessed on a scale of 1 to 10, currently averages at 3.38. This reflects the current state of the airline industry in India, which has remained stagnant over the past decade.

Measurement Model Estimation

To interpret the results of this analysis, we first examine the measurement model which is followed by the examination of structural model and serial mediation testing.

Table no. 2 shows the results of measurement model estimation to examine how well each indicator is loaded onto its respective construct. As a rule of thumb factor loadings should be above 0.708, and in our case, all the variables met that threshold, indicating that they are reliable. We also evaluated internal consistency using Cronbach's alpha, which was above 0.6, and composite reliability, which exceeded 0.7. Both values are above the minimum thresholds showing that the variables are consistent and reliable. Additionally, each construct had an average variance extracted (AVE) greater than 0.50, confirming that they captured enough information from their indicators. Lastly, convergent validity was confirmed because the composite reliability (CR) was higher than the AVE, concluding that our constructs are reliable and valid.

Table 2: Measurement Model Estimation

Constructs	Standardized Factor Loading	Composite Reliability	Average Variance Extracted	Cronbach's Alpha
Customer Satisfaction		0.886	0.795	0.748
Sentiment	0.854			
value_for_money	0.928			

Operational Efficiency		0.906	0.828	0.793
ground_service	0.915			
staff_service	0.905			
Service Quality		0.926	0.806	0.88
entertainment	0.894			
food_quality	0.894			
seating_comfort	0.904			

In the next step, the discriminant validity is checked using the Fornell-Larcker criterion. Table no. 3 shows the results of this criterion. As shown in the table, the square root of AVE of the constructs were higher than their correlations with other variables indicating that our data is free of discriminant validity.

Table 3: Fornell-Larcker Criterion

	Customer Satisfaction	Operational Efficiency	Service Quality	overall_rating
Customer Satisfaction	**0.892**			
Operational Efficiency	0.8	**0.91**		
Service Quality	0.819	0.803	**0.898**	
overall_rating	0.881	0.843	0.86	**1**

Structural Model Estimation

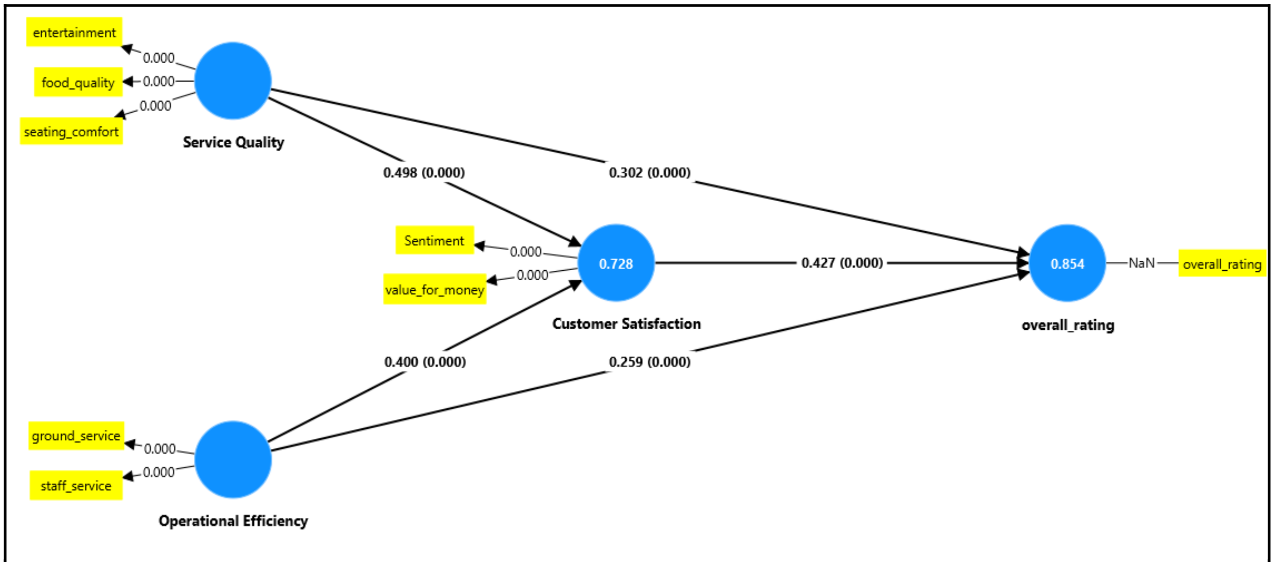


Diagram 4: Model Estimation Results

Source: Ringle, C. M., Wende, S., and Becker, J.-M. 2024. "SmartPLS 4." Bönningstedt: SmartPLS, https://www.smartpls.com.

Direct Effect

To estimate the structural model as shown in table no 4 using SmartPLS software's PLS-SEM Algorithm and bootstrapping procedure with 5000 subsamples. The direct, indirect and mediating effect is measured as shown in the figure no... above. The results of the direct effect are shown in table no. 4.

Table 4: Direct Relationship

Hypothesis	**Path Diagram**	**Std β - Value**	***Standard Deviation***	**t-Statistic**	**p-values**
H1	Service Quality -> Customer Satisfaction	0.498	0.017	28.594	0.000
H2	Service Quality -> overall_rating	0.302	0.016	18.862	0.000
H3	Operational Efficiency -> Customer Satisfaction	0.400	0.017	22.958	0.000
H4	Operational Efficiency -> overall_rating	0.259	0.014	19.053	0.000
H5	Customer Satisfaction -> overall_rating	0.427	0.014	30.796	0.000

Hypothesis 1 tests whether Customer Satisfaction is explained by Service Quality. According to the results of path analysis in table no. 4 Service Quality (path coefficient =0.498, t-value = 28.594, and p-value<0.05) significantly affects Customer Satisfaction. Hypothesis 2 tests whether Overall Rating is explained by Service Quality. Service Quality (path coefficient = 0.302, t-value = 18.862, and p-value<0.05) significantly affects Overall Rating. Hypothesis 3 tests whether Customer Satisfaction is explained by Operational Efficiency. Operation Efficiency (path coefficient =0.4, t-value = 22.958, and p-value<0.05) significantly affects Customer Satisfaction. Hypothesis 4 tests whether Overall Rating is explained by Operational Efficiency. Operation Efficiency (path coefficient =.0.259, t-value =19.053, and p-value<0.05) significantly affects the Overall Rating. Hypothesis 5 tests whether Overall Rating is explained by Customer Satisfaction. Customer Satisfaction (path coefficient =0.427, t-value = 0.014, and p-value<0.05) significantly affects the Overall Rating.

Test for Mediating Effects

The model further includes serial mediation between service quality and overall rating, and between operational efficiency and overall rating. The mediating variable in both cases is customer satisfaction. This mediation is tested with path coefficients and the significance values of the direct and indirect paths. As shown in table no. 5 the direct relationship between all the variables is significant. Hence, the basic condition for serial mediation analysis, i.e. having a significant relationship, is fulfilled.

Table 5: Result of Direct & Indirect Effects

Relationship Between	Type of Effect	Path Coefficient	T Stats	P value	Remark
Service Quality -> overall_rating	Total Effect	0.515	30.82	0.000	Total effect is significant
Service Quality -> Customer Satisfaction -> overall_rating	Indirect Effect	0.212	22.115	0.000	Serial mediation exist
Service Quality -> overall_rating	Direct Effect	0.302	18.862	0.000	Direct effect is significant
Operational Efficiency -> overall_rating	Total Effect	0.429	26.541	0.000	Total effect is significant
Operational Efficiency-> Customer Satisfaction -> overall_rating	Indirect Effect	0.171	17.906	0.000	Serial mediation exist
Operational Efficiency -> overall_rating	Direct Effect	0.259	19.053	0.000	Direct effect is significant

According to table no 5 , the total effect ($\beta = 0.515$, $t = 30.82$, $p<0.05$) of service quality on overall rating is significant in the presence of a mediating variable. The indirect effect ($\beta = 0.212$, $t= 22.115$, $p<0.05$) of service quality to overall rating through customer satisfaction serial mediation is significant. The direct effect ($\beta = 0.302$, $t= 18.86$, $p<0.05$) of service quality on overall rating is significant. Therefore, in the presence of the mediator customer satisfaction, the direct effect is sustained and significant indicating that the relationship is partially mediated. It can be concluded that Service quality and Overall Rating has a partial mediated relationship in the presence of Customer Satisfaction. The same can be interpreted for Operational Efficiency and Overall Rating.

After obtaining satisfactory results from the measurement model, we further assessed the standard assessment R^2. The R^2 score is commonly known as the coefficient of determination. Table No. 6 shows the R^2 values for the dependent variable.

Table 6: R^2 Score

	R-square
Customer Satisfaction	0.728
overall_rating	0.854

The R^2 score for the dependent variable customer satisfaction is 0.728. This means that around 72.8% variation in customer satisfaction is explained jointly by the service quality and operational efficiency. Similarly, the R^2 score for overall rating is 0.854, indicating that around 85.4% of variation in overall rating is explained by the other three variables, service quality, operational efficiency and customer satisfaction together.

Conclusion

The sentiment analysis of airline reviews has provided crucial insights into customer satisfaction within the Indian airline industry. It highlights that airlines such as Vistara and SpiceJet excel in overall service, seating comfort, and staff behavior, particularly in premium classes. In contrast, Air India and Indigo face challenges in Economy Class, where passengers have expressed dissatisfaction with food quality, staff responsiveness, and seating arrangements. Addressing these issues could significantly enhance the overall customer experience and improve ratings.

Following the sentiment analysis, regression results reveal significant relationships between service quality, operational efficiency, customer satisfaction, and overall airline ratings. Key findings indicate that both service quality and operational efficiency have a substantial impact on overall ratings with customer satisfaction being the mediating variable.

In conclusion, this analysis highlights the importance of continuously monitoring customer feedback through sentiment analysis to make data-driven decisions. Airlines can use these insights to implement service improvements, enhance customer satisfaction, and ultimately maintain competitiveness in a highly dynamic and customer-centric industry.

References

Banerji, D., Saha, V., Singh, N., & Srivastava, R. (2023). What are the most important consumer decision factors when choosing an airline? An emerging economy perspective. Asia Pacific Journal of Marketing and Logistics, 35(1), 174-197.

Bunchongchit, K., & Wattanacharoensil, W. (2021). Data analytics of Skytrax's airport review and ratings: Views of airport quality by passengers types. Research in Transportation Business & Management, 41, 100688. https://doi.org/10.1016/j.rtbm.2021.100688

Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. Journal of the Academy of Marketing Science, 40(3), 414–433.

Hair, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2013). A primer on partial least squares structural equation modeling (PLS-SEM). SAGE Publications.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. European Business Review, 31(1), 2–24.

Ozkul, A., et al. (2020). Factors affecting customer satisfaction in the airline industry: A systematic review. Journal of Air Transport Management, 89(1), 101-112.

BairesDev. (2024). 8 Best Python Sentiment Analysis Libraries. Retrieved from https://www.bairesdev.com/blog/best-python-sentiment-analysis-libraries

Labellerr. (2024). Top 7 NLP Libraries For NLP Development In 2024. Retrieved from https://www.labellerr.com/blog/top-7-nlp-libraries-for-nlp-development/