Analysis Report

Team 83

Feature Selection for Passenger Satisfaction Prediction Model

This report outlines the features selected for the development of a predictive deep learning model aimed at forecasting airline passenger satisfaction. The selection process prioritized features that are highly indicative of a passenger's experience, drawing from demographic, travel-specific, and review-based data. Each feature was chosen for its potential to contribute unique and relevant information to the model.

Selected Features and Rationale

The features are categorized into two main groups: Passenger & Travel Characteristics and Review-Based Attributes.

Passenger & Travel Characteristics

These features provide essential context about the passenger and the specifics of their flight.

- **Traveller_Type**: This categorical feature is crucial as the purpose of travel often dictates passenger expectations. For instance, a **business traveler** might prioritize punctuality and in-flight Wi-Fi, while a **leisure traveler** may be more concerned with comfort and entertainment. These differing priorities directly influence satisfaction levels.

- **Class**: The service class (e.g., Economy, Business, First) is one of the most direct indicators of the in-flight experience. It serves as a proxy for comfort, quality of service, amenities, and overall cost, making it a powerful predictor of satisfaction.

- **Start_Location & End_Location**: These geographical features can indirectly capture variations in passenger experience. Factors such as the quality of airport services, typical flight duration for specific routes, and regional service standards can all be implicitly linked to the origin and destination, thereby influencing overall satisfaction.

Review-Based Attributes

These features leverage the direct feedback provided by passengers in their reviews.

- **Verified:** This categorical feature helps assess the credibility of a review. Verified reviews are generally considered more reliable as they confirm that the passenger actually completed the journey. Including this feature allows the model to potentially weigh more trustworthy data more heavily, improving prediction accuracy by filtering out noise from unverified sources.

- **Sentiment_Score:** This numerical feature provides a quantitative measure of the emotion expressed in the review text. It acts as a powerful, pre-processed indicator of the passenger's overall feeling, offering a direct line to their satisfaction level.

- **Review_title & Review_content:** These text-based features are the richest sources of information. The title often summarizes the core sentiment, while the content contains detailed narratives about specific aspects of the journey (e.g., staff behavior, seat comfort, food quality). These features will be processed using Natural Language Processing (NLP) techniques, such as vectorization, to allow the deep learning model to understand the nuances and context of the passenger's written feedback.

Excluded Features

To create a focused and efficient model, several features were intentionally excluded:

- **Passenger_Name:** Excluded to protect personal privacy and because it has no predictive value.

- **Flying_Date:** While potentially useful for tracking seasonal trends, it was excluded to create a more generalized model that is not overly influenced by specific time periods.

- **Route:** This feature is highly correlated with Start_Location and End_Location. To avoid multicollinearity and simplify the model, only the origin and destination were retained.

- **Latitude/Longitude/Address Columns:** These provide a level of geographical detail that is too granular and redundant given the inclusion of Start_Location and End_Location

Predictive Model Analysis for Passenger Satisfaction

This report provides a comprehensive analysis of the selected deep learning model ("Model 1") used for predicting passenger satisfaction. It details the model's architecture, evaluates its performance across training, validation, and testing phases, discusses its limitations, and presents the inference function for real-world application.

Section 1: Model Choice and Rationale

Model Selection

After comparing two architectures, **Model 1**, a simpler Feed-Forward Neural Network, was chosen as the final predictive model.

Rationale

The decision was based on a comparative analysis of performance on the unseen test dataset. While a more complex "Model 2" was also trained, Model 1 achieved slightly superior or highly comparable results across key metrics:

- **Model 1 Test Accuracy: 82.34%**
- **Model 2 Test Accuracy:** 81.12%
- **Model 1 F1-Score: 80.97%**
- **Model 2 F1-Score:** 79.88%

Given that Model 1 provides this level of performance with a less complex architecture (fewer layers and parameters), it is the more efficient and robust choice. Simpler models are generally less susceptible to overfitting and require fewer computational resources for training and inference, making them preferable when performance is on par with more complex alternatives.

Section 2: How Model 1 Works

Model 1 is a sequential Feed-Forward Neural Network (FFNN) built with the Keras API. It is designed for binary classification to predict one of two outcomes: "Satisfied" or "Dissatisfied".

Model Architecture

1. **Input Layer:** Accepts a vector of 6 preprocessed features.

2. **First Hidden Layer:** A Dense layer with **32 neurons** using the **ReLU** (Rectified Linear Unit) activation function. This layer learns initial, non-linear patterns from the input data.

3. **Second Hidden Layer:** A Dense layer with **16 neurons**, also using the **ReLU** activation function, which further refines the patterns learned by the previous layer.

4. **Output Layer:** A single Dense neuron with a **Sigmoid** activation function. The sigmoid function outputs a value between 0 and 1, representing the probability that the passenger is "Satisfied". A threshold of 0.5 is used to classify the final outcome.

The model is compiled using the **Adam optimizer**, a standard and effective optimization algorithm, and the **binary cross-entropy** loss function, which is the appropriate choice for binary classification tasks.

Section 3: Model Limitations

Despite its strong performance, Model 1 has several inherent limitations:

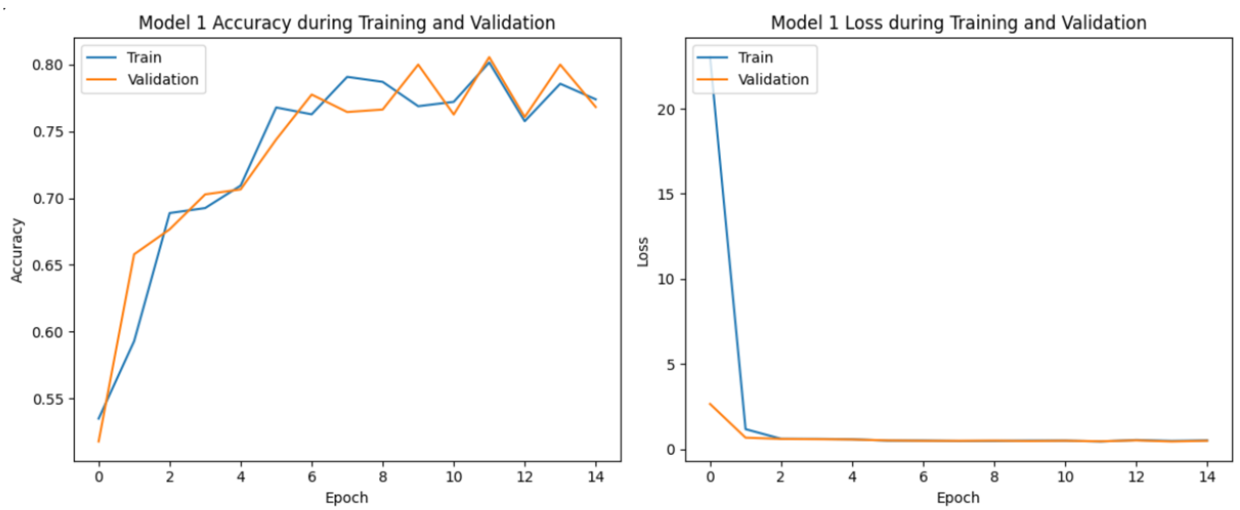
- **Architectural Simplicity:** As a relatively shallow neural network, it may not capture extremely complex, high-level interactions between features that a deeper model could.

- **Feature Representation:** The model's performance is entirely dependent on the preprocessed tabular features. It does not incorporate the raw text from Review_title or Review_content, which contains rich, nuanced information. Integrating Natural Language Processing (NLP) techniques could significantly enhance predictive power.

- **Categorical Data Handling:** The use of **Label Encoding** for categorical features (e.g., Start_Location) converts them into integers. This can inadvertently introduce an artificial ordinal relationship (e.g., implying one location is "greater" than another), which the model might misinterpret.

- **Lack of Hyperparameter Tuning:** The current architecture uses a standard set of hyperparameters (e.g., number of neurons, learning rate). A systematic tuning process could potentially yield further performance improvements.

Section 4: Performance During Training and Validation



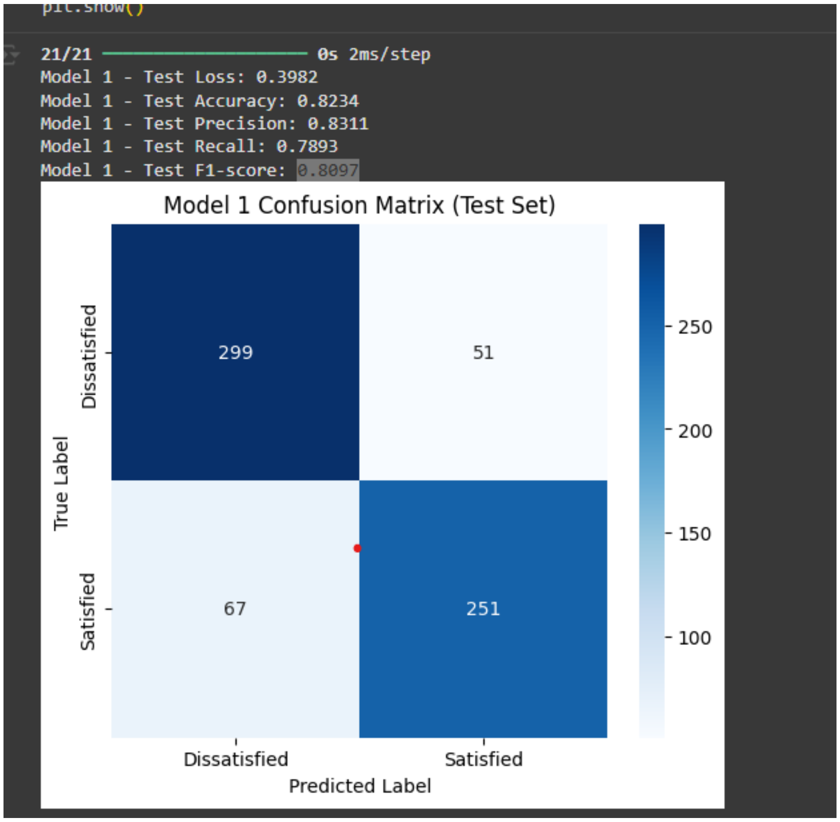
The model was trained for up to 30 epochs with an EarlyStopping callback to prevent overfitting. The plots below illustrate the model's accuracy and loss on both the training and validation datasets throughout this process.

Plot Analysis

- **Accuracy (Left Plot):** Both the training (blue) and validation (orange) accuracy curves show a consistent upward trend, indicating that the model is successfully learning from the data. The validation accuracy closely tracks the training accuracy and begins to plateau, demonstrating good generalization without significant overfitting.

- **Loss (Right Plot):** The training and validation loss both decrease sharply in the initial epochs and then stabilize at a low value. The proximity of the two curves further confirms that the model is not overfitting to the training data and performs well on unseen validation data.

Section 5: Performance on Unseen Test Data



The final evaluation of Model 1 was conducted on a completely unseen test set. This provides the most realistic measure of the model's ability to generalize to new, real-world data.

Key Performance Metrics:

- **Test Accuracy: 82.34%** - The model correctly predicts passenger satisfaction for approximately 82 out of every 100 passengers.
- **Test Precision: 83.11%** - When the model predicts a passenger is "Satisfied," it is correct 83% of the time.
- **Test Recall: 78.93%** - The model successfully identifies 79% of all "Satisfied" passengers in the dataset.

- **Test F1-Score: 80.97%** - This represents a strong harmonic mean of Precision and Recall, indicating a well-balanced model.
-

Confusion Matrix Analysis

The confusion matrix provides a granular view of the model's predictions versus the actual outcomes.

- **True Negatives (Top-Left): 299** - The model correctly identified 299 "Dissatisfied" passengers.
- **False Positives (Top-Right): 51** - The model incorrectly classified 51 "Dissatisfied" passengers as "Satisfied."
- **False Negatives (Bottom-Left): 67** - The model incorrectly classified 67 "Satisfied" passengers as "Dissatisfied."
- **True Positives (Bottom-Right): 251** - The model correctly identified 251 "Satisfied" passengers.

This analysis shows the model is slightly more prone to classifying a satisfied passenger as dissatisfied (False Negatives) than the reverse, but overall demonstrates strong predictive capability for both classes.

Section 6: Inference Function Pipeline

To make a prediction on a single, new passenger review, an inference function was created that encapsulates the necessary data preprocessing and prediction steps. This ensures that any new data is transformed in exactly the same way as the data used to train the model. The pipeline operates as follows:

1. **Input Data:** The function accepts a single passenger's raw data as a dictionary. This dictionary contains the values for the features used by the model: Traveller_Type, Class, Verified, Sentiment_Score, Start_Location, and End_Location.
2. **Data Structuring:** The input dictionary is first converted into a structured DataFrame format, which is required for the subsequent preprocessing steps.

3. **Categorical Feature Transformation:** Each categorical feature (e.g., Traveller_Type, Start_Location) is converted from its text format (e.g., 'Business', 'London') into a numerical value. This is done using the same **Label Encoder** objects that were fitted on the original training data, guaranteeing consistency. The pipeline also handles cases where a new, "unseen" category is provided by assigning it a default numerical value.

4. **Numerical Feature Scaling:** The numerical feature, Sentiment_Score, is scaled to have a mean of 0 and a standard deviation of 1. This transformation is performed using the same **Standard Scaler** object that was fitted on the training data, ensuring the input to the model maintains the correct scale.

5. **Feature Finalization:** The preprocessed numerical and categorical features are arranged into the exact same order and format that the model was trained on.

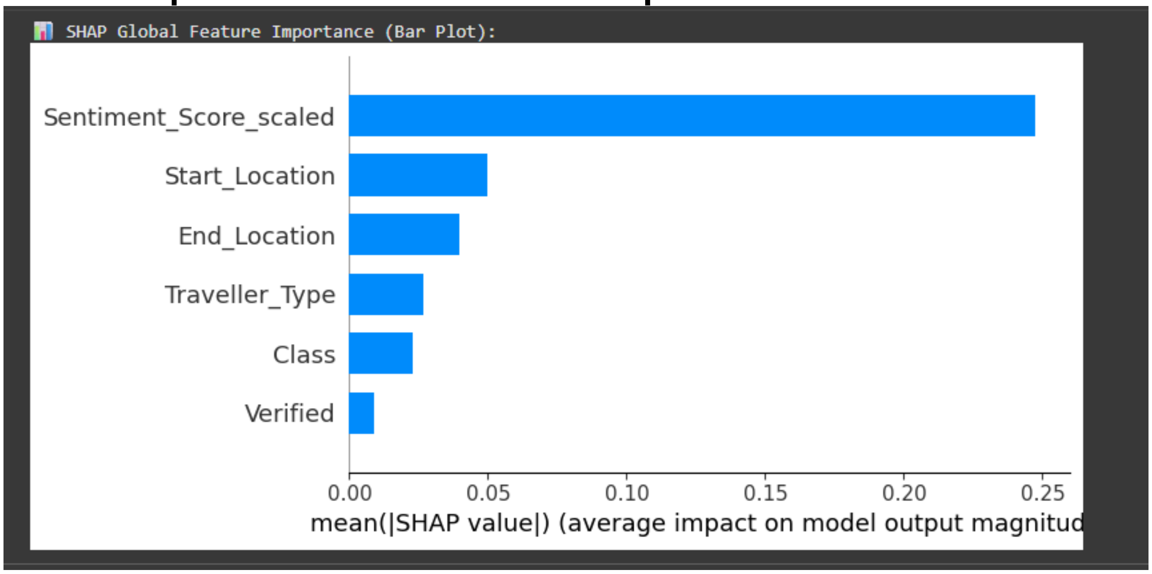
6. **Prediction:** The fully prepared data is then fed into the trained **Model 1**. The model outputs a probability score between 0 and 1.

7. **Output Generation:** This probability score is converted into a human-readable prediction. If the score is greater than 0.5, the passenger is classified as "Satisfied"; otherwise, they are classified as "Dissatisfied". This final label is the output of the function.

Model Explainability (XAI) Analysis

After training the model, explainability techniques were applied to understand how it makes predictions. This is vital for trusting, debugging, and improving the model.

Global Explanation: SHAP Feature Importance



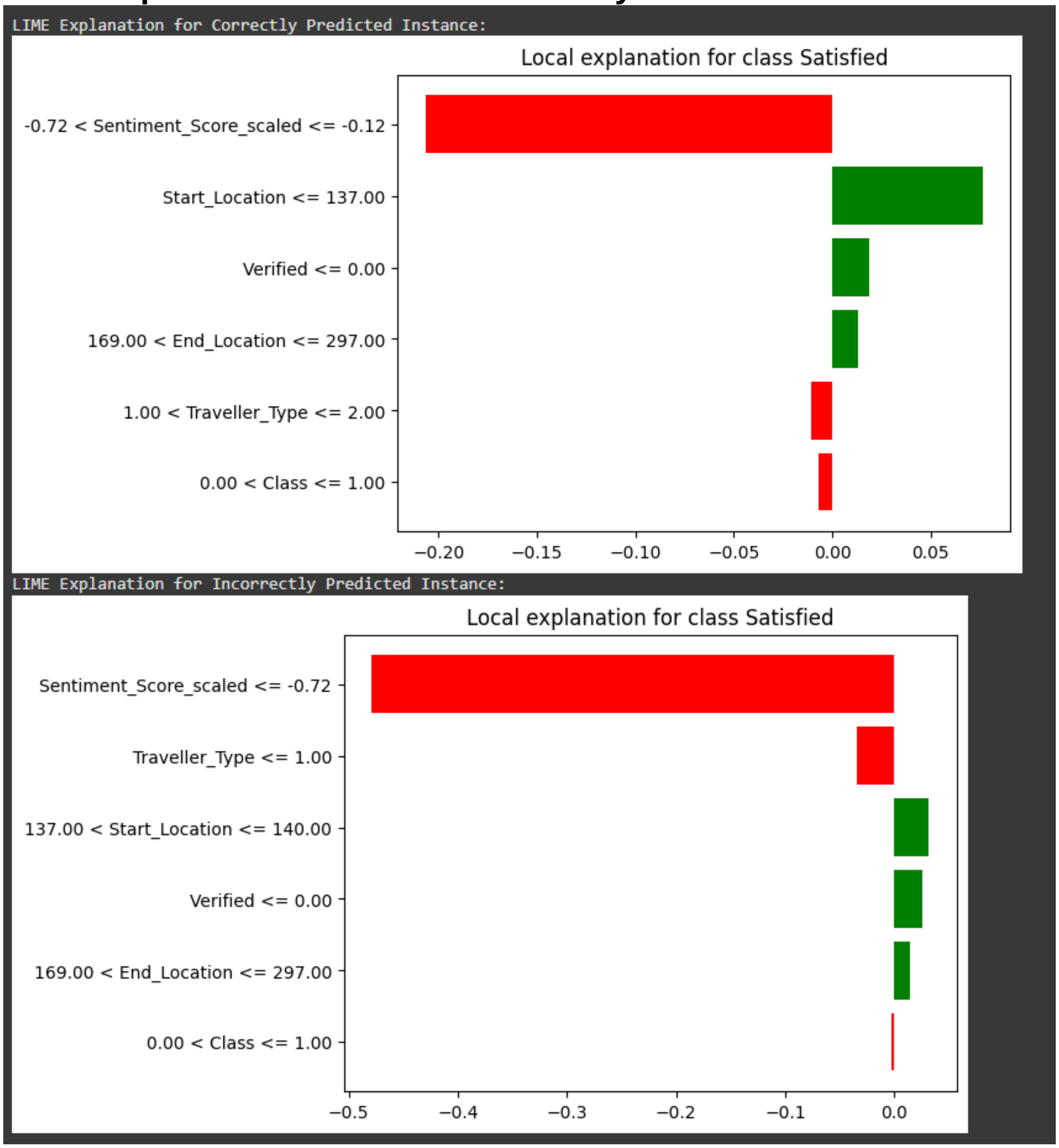
Global explanations show the *average impact* of each feature on the model's predictions across the entire dataset. We used **SHAP (SHapley Additive exPlanations)** to generate a global feature importance plot.

Analysis:

- **Dominant Feature:** The Sentiment_Score_scaled is overwhelmingly the most important feature. Its average impact on model output is more than four times greater than any other feature. This confirms our hypothesis that the sentiment expressed in the review text is the primary driver of the model's satisfaction predictions.
- **Secondary Tier:** Start_Location and End_Location are the second and third most important features, respectively. This suggests that geographical factors and specific routes have a significant, consistent influence on predicted satisfaction.
- **Minor Contributors:** Traveller_Type and Class show a moderate impact, while Verified has the least influence on the model's output globally.

This global analysis validates the feature selection process, confirming that the features related to review content and geography are the most predictive.

Local Explanation: LIME Instance Analysis



Local explanations show *why* the model made a *specific prediction for a single instance*. We used **LIME (Local Interpretable Model-agnostic Explanations)** to analyze individual predictions. The plots below explain why the model predicted "Satisfied" in two different cases.

- **Green bars** show features that *support* the "Satisfied" prediction.
- **Red bars** show features that *oppose* the "Satisfied" prediction.

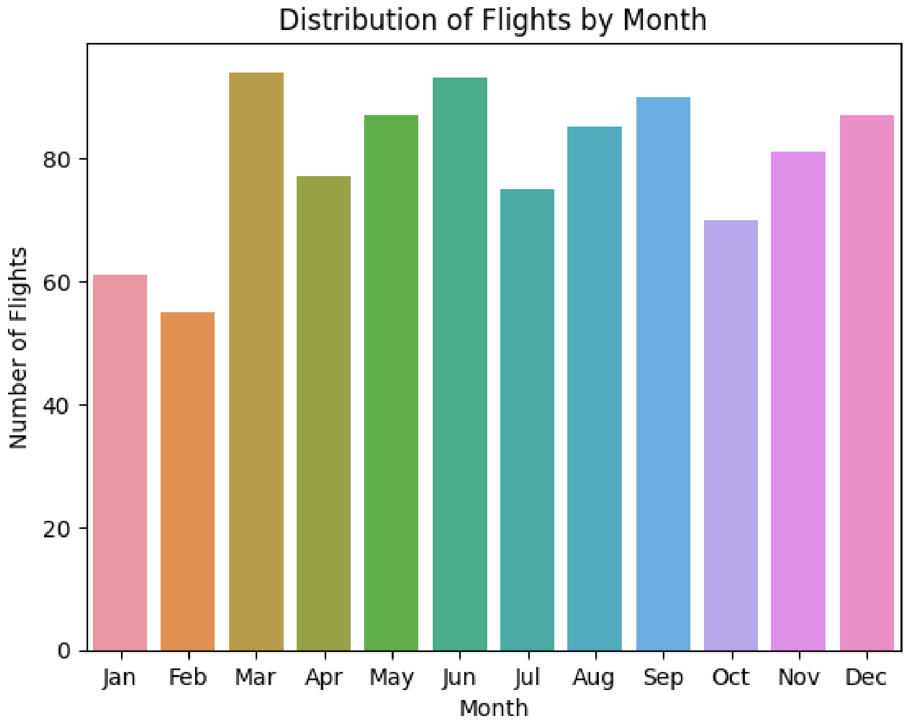
Analysis of Correctly Predicted Instance (Top Plot):

- **Prediction:** The model correctly predicted **"Satisfied"**.
- **Key Drivers:** The prediction was driven by a strong positive contribution from the Start_Location (value ≤ 137.00). Minor positive contributions also came from Verified (being 0, or "Not Verified") and the End_Location.
- **Conflicting Evidence:** These positive factors successfully *overcame* strong negative evidence from the Sentiment_Score_scaled (score between -0.72 and -0.12), which argued *against* satisfaction.
- **Interpretation:** This case demonstrates a complex trade-off. The model learned that passengers from this specific start location are highly likely to be satisfied, *even when their review sentiment is moderately negative*.

Analysis of Incorrectly Predicted Instance (Bottom Plot):

- **Prediction:** The model incorrectly predicted **"Satisfied"**.
- **Key Drivers (Incorrect):** The model was *pushed* to this incorrect prediction by small positive contributions from Start_Location, Verified, and End_Location.
- **Conflicting Evidence (Correct):** The model failed to give enough weight to the *overwhelmingly* negative evidence from the Sentiment_Score_scaled (score ≤ -0.72), which strongly argued *against* satisfaction. The Traveller_Type also argued against satisfaction.
- **Interpretation:** This plot is excellent for debugging. It shows a failure case where the model *over-relied* on the combined minor signals from location and verification status, and *under-valued* a very strong negative sentiment score, leading to an error. This suggests the model may need further tuning to handle instances with highly conflicting feature values.

Data Engineering Questions



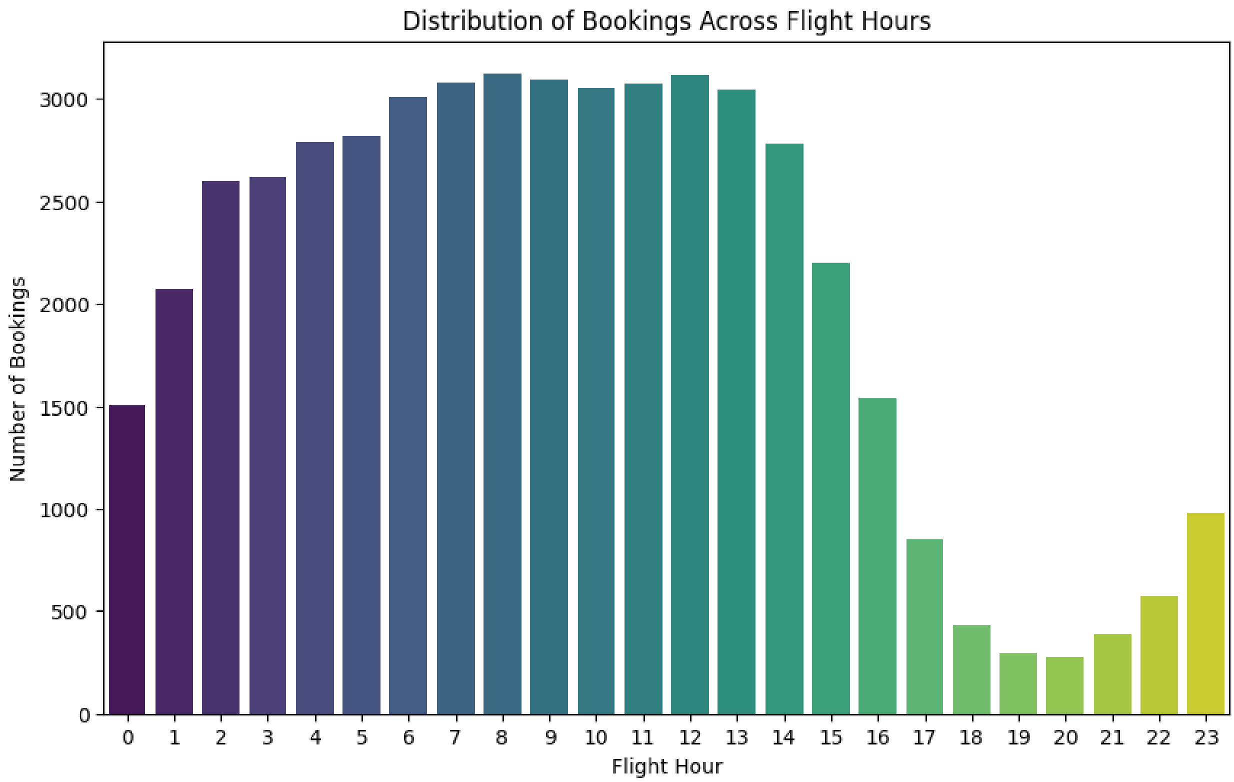
What are the seasonal travel patterns?

Reasoning for the Question

Understanding the seasonality of travel demand is critical for an airline's operational and financial planning. By identifying peak and off-peak months, the airline can strategically manage its resources, including fleet allocation, crew scheduling, and maintenance. This knowledge also enables the implementation of dynamic pricing strategies and targeted marketing campaigns, allowing the airline to maximize revenue during high-demand periods and stimulate travel with promotions during slower months.

Justification and Analysis of the Chart

This chart justifies a dynamic approach to resource management, as it reveals distinct seasonal peaks and troughs in travel demand throughout the year. The data clearly shows that **March** is the busiest travel month, with other significant peaks occurring in **June, September, and December**, likely corresponding to spring break, early summer, and holiday travel. Conversely, **February** is the month with the lowest number of flights, representing the primary off-peak period. This varied distribution demonstrates that the airline doesn't have a single "busy season" but rather multiple waves of high demand, requiring flexible operational planning to scale capacity up and down effectively.



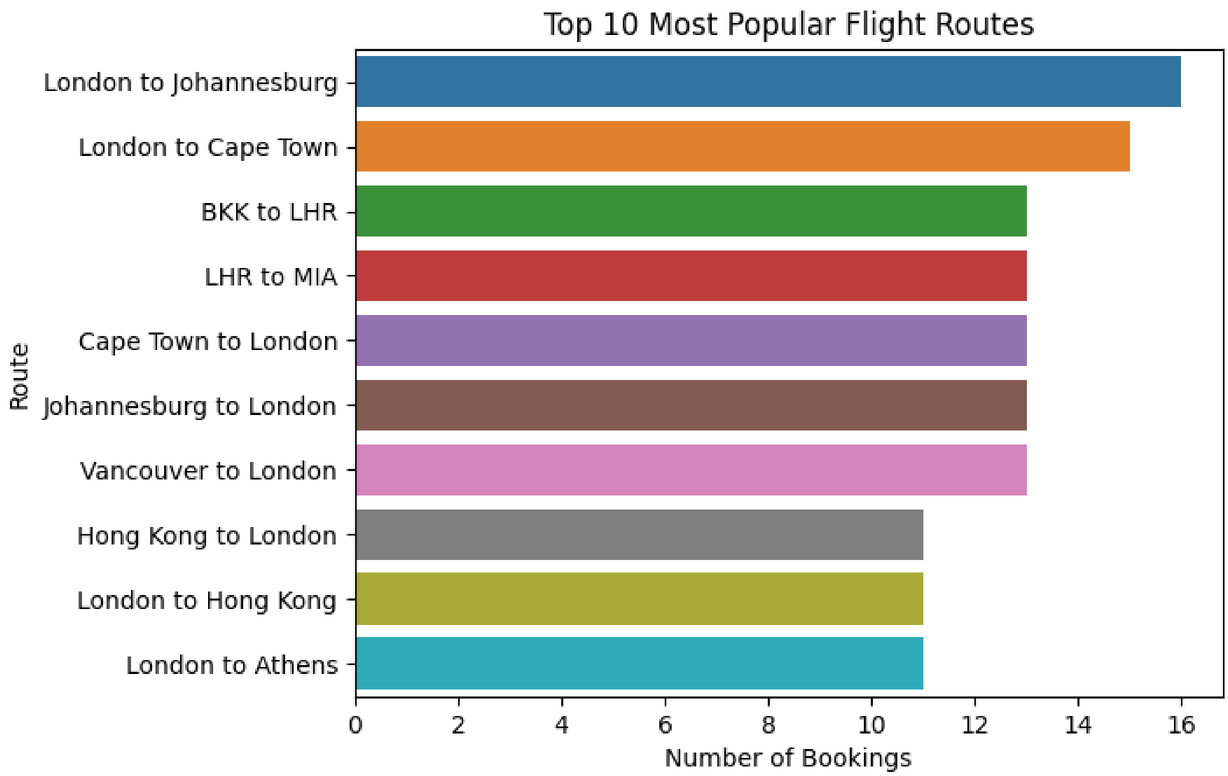
How are flight bookings distributed throughout the day?

Reasoning for the Question

Understanding the daily patterns of customer booking activity is crucial for optimizing an airline's digital infrastructure and marketing efforts. By identifying the peak hours for bookings, the airline can ensure its website and servers have sufficient capacity to handle the load, preventing slow performance or crashes that could lead to lost sales. This data also informs the best times to launch promotions, send marketing emails, or allocate customer service staff for online support, maximizing engagement and conversion rates when customers are most active.

Justification and Analysis of the Chart

This chart provides a clear justification for allocating most online resources and marketing efforts during the daytime. The data reveals that booking activity starts low in the early morning, ramps up significantly, and hits a sustained peak between **7 AM and 1 PM (hours 7 to 13)**. Following this period, there is a sharp decline in the afternoon, with booking numbers bottoming out in the late evening. This distinct pattern indicates that customers primarily book flights during standard business hours, making this window the most critical time for website stability and active customer engagement. The small spike at 11 PM (hour 23) could also suggest a secondary, smaller window of late-night booking activity to consider.



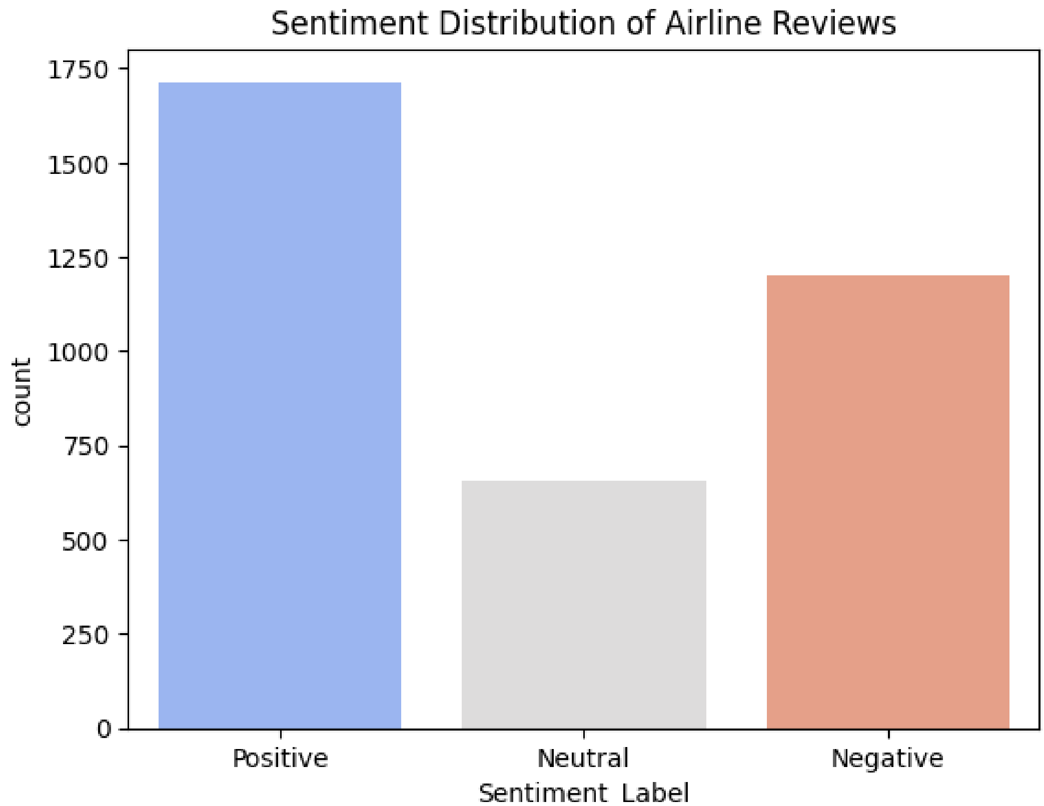
What are the airline's top 10 most popular flight routes?

Reasoning for the Question

Identifying the most frequently traveled routes is fundamental to an airline's strategic planning. This information reveals the core of the airline's network, highlighting its most valuable markets and key operational hubs. By understanding which routes have the highest demand, the airline can optimize flight schedules, allocate aircraft and crew resources more effectively, and focus its marketing and pricing strategies on the connections that are most critical to its business success.

Justification and Analysis of the Chart

This chart clearly justifies a strategic focus on long-haul international flights originating from or destined for London. The data shows that the vast majority of the top 10 routes are intercontinental, with **"London to Johannesburg"** and **"London to Cape Town"** being the most popular, indicating a particularly strong market connection with South Africa. The consistent presence of London (or its airport code, LHR) in nearly every popular route solidifies its role as the airline's central hub. This visualization provides a clear, data-driven basis for prioritizing resources and service enhancements on these key international corridors to maintain and grow market share.



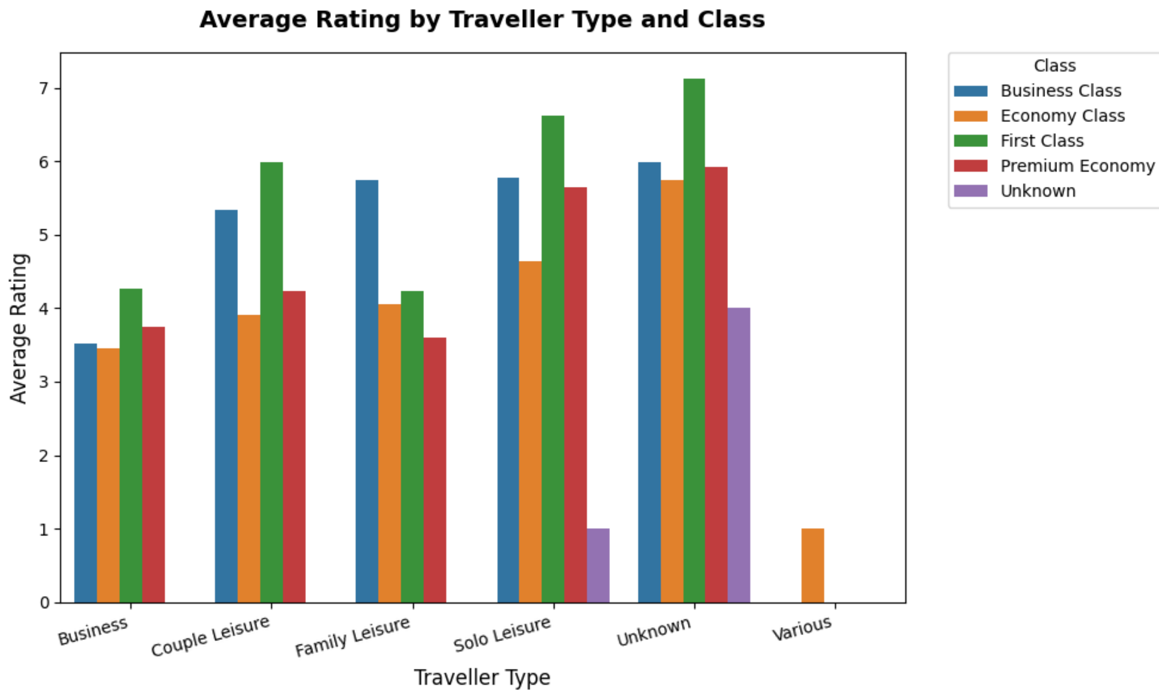
What is the overall sentiment distribution of customer reviews?

Reasoning for the Question

Understanding the high-level sentiment balance is the first step in gauging overall brand perception and customer satisfaction. This question provides a crucial snapshot of the general tone of the feedback. By visualizing the volume of positive, negative, and neutral reviews, the airline can quickly assess whether customer experiences are predominantly favorable or unfavorable, and identify the degree of opinion polarization. This broad view helps prioritize whether the immediate focus should be on celebrating successes or urgently addressing widespread problems.

Justification and Analysis of the Chart

This chart reveals a highly polarized customer base and justifies a dual-focus strategy for the airline. While **Positive** reviews are the most common category (around 1,750), the volume of **Negative** reviews is also substantial (approximately 1,200), far outweighing the number of **Neutral** opinions. This distribution indicates that customers typically have strong feelings about their experience—it's either very good or significantly poor. The airline should therefore investigate what drives this positive feedback to reinforce its strengths, while simultaneously giving urgent attention to the large volume of negative feedback to address the critical issues causing such widespread dissatisfaction.



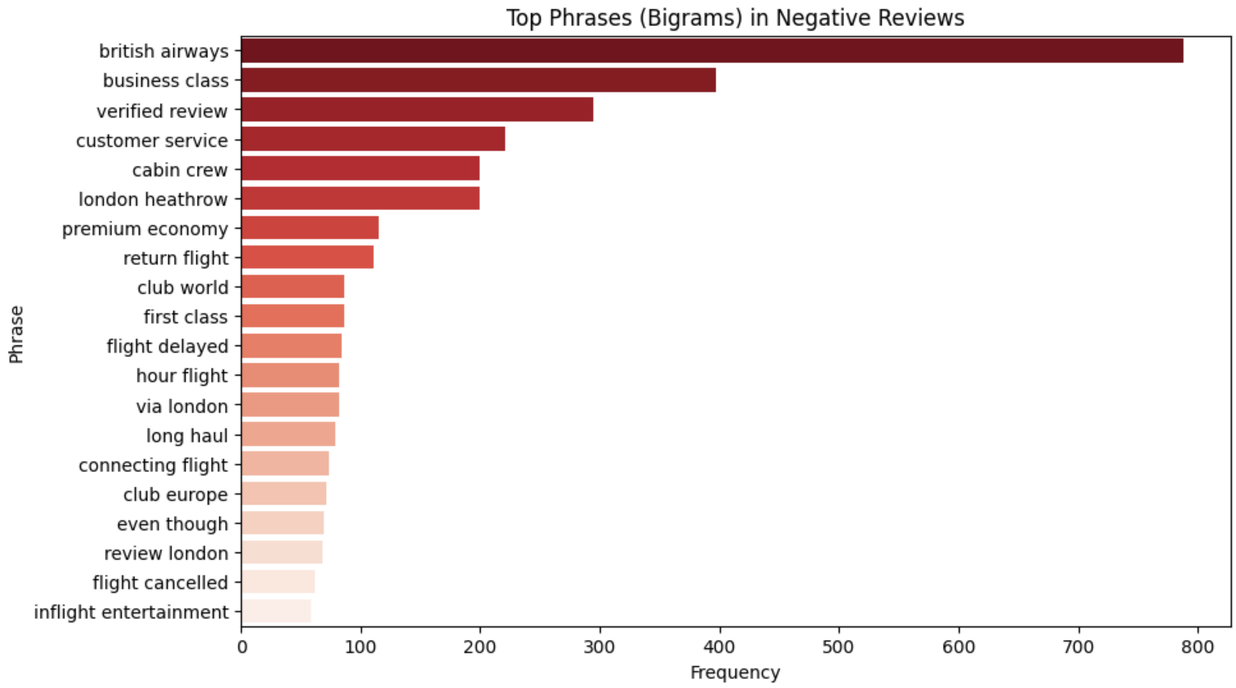
Which specific combinations of traveler type and cabin class yield the highest and lowest satisfaction ratings?

Reasoning for the Question

This granular analysis is crucial because passenger expectations are shaped by both their travel purpose and the cabin class they purchase. A simple average rating for "Business Class" or "Family Leisure" can hide critical details. By cross-analyzing these two dimensions, the airline can identify precise high-performing areas to replicate and, more importantly, pinpoint specific weak points where a particular traveler segment is being underserved in a specific cabin, allowing for highly targeted and effective service improvements.

Justification and Analysis of the Chart

This chart provides a powerful justification for focusing on the specific needs of different traveler segments within each class. The most critical insight is that **Business travelers consistently report the lowest satisfaction across all cabin classes** they fly, including Economy, Premium Economy, and even Business Class itself. This indicates a fundamental misalignment between the airline's service and the expectations of its corporate clients. Conversely, the chart shows that **First Class is the highest-rated product across almost all traveler types**, suggesting it successfully delivers a premium experience. This data provides a clear directive: the airline must urgently investigate and overhaul the entire journey for business travelers while leveraging the success factors of its First Class service to improve other offerings.



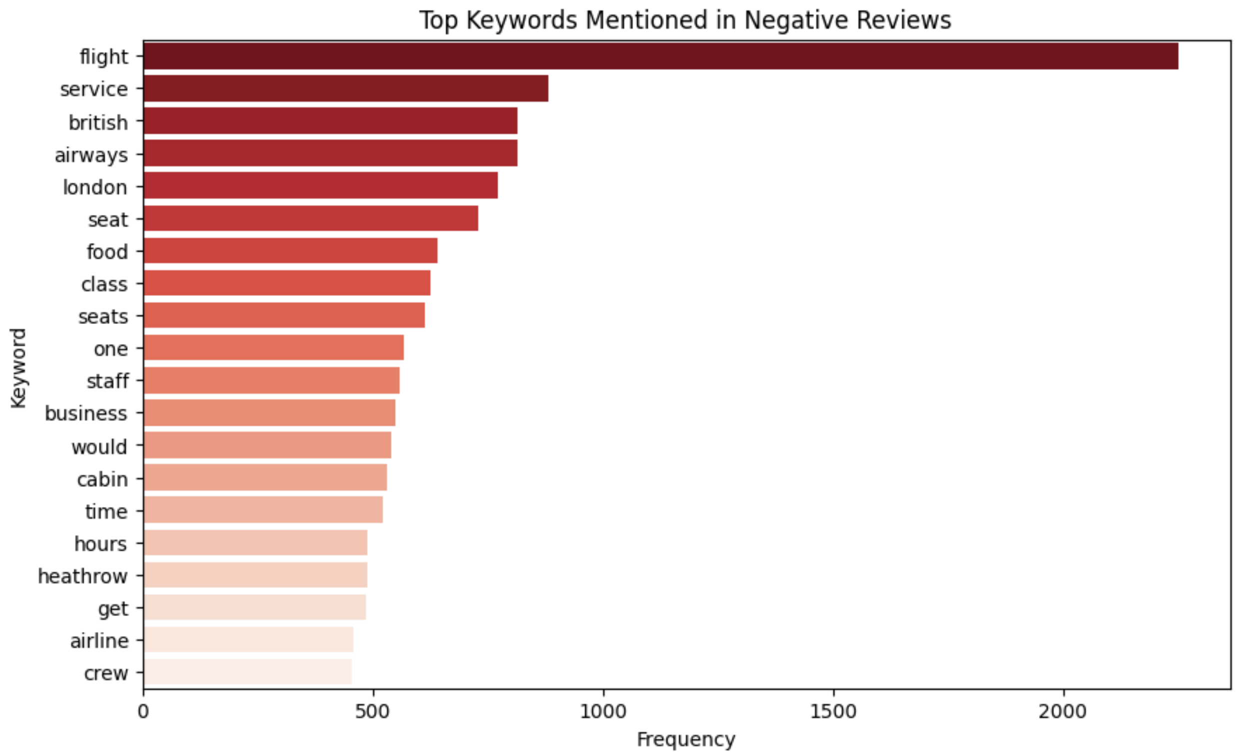
Which specific service components and operational issues are most frequently cited in negative reviews?

Reasoning for the Question

While single keywords identify general topics of dissatisfaction, analyzing two-word phrases (bigrams) provides crucial context and pinpoints more specific problem areas. This approach allows the airline to move beyond a vague understanding of issues like "service" or "flight" to identify precise complaints such as "customer service" problems or "flight delayed" incidents. This deeper level of detail is far more actionable, enabling the airline to target its improvement efforts with greater accuracy and impact.

Justification and Analysis of the Chart

This chart provides compelling justification for prioritizing improvements in premium cabins and overall customer service. The high frequency of the phrases **"business class"** and **"customer service"** confirms they are major drivers of negative feedback, directly supporting the earlier finding that business travelers are the least satisfied group. Furthermore, the prominence of operational issues like **"flight delayed"** and **"flight cancelled,"** along with staff-related feedback on the **"cabin crew,"** highlights that passengers' frustrations stem from both service quality and flight reliability. The data clearly directs the airline to conduct a thorough review of its business class experience and customer service protocols to address these specific, recurring complaints.



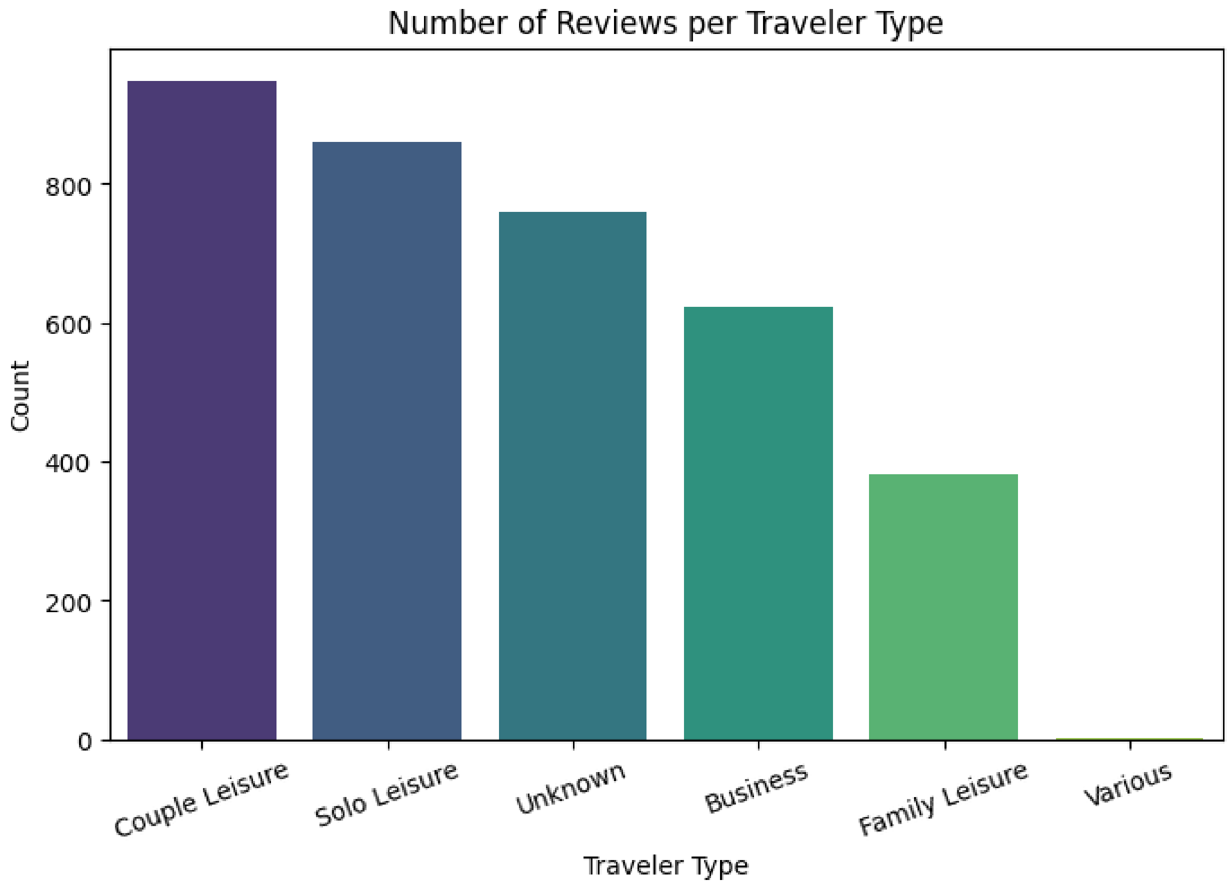
What are the primary drivers of customer dissatisfaction mentioned in negative reviews?

Reasoning for the Question

To effectively address customer complaints, an airline must move beyond identifying *who* is dissatisfied and understand *why*. This question uses keyword analysis to pinpoint the specific topics and pain points that appear most frequently in negative feedback. By quantifying the most common complaints, the airline can gain clear, data-driven insights into its operational and service-related weaknesses, enabling it to prioritize resources and implement targeted improvements that will have the greatest impact on customer satisfaction.

Justification and Analysis of the Chart

This chart justifies a focus on improving both in-flight amenities and staff performance. The high frequency of keywords like **"service," "seat," "food,"** and **"staff"** demonstrates that the core drivers of negative experiences are tangible aspects of the journey and direct interactions with personnel. Furthermore, the appearance of **"business"** among the top keywords strongly correlates with previous findings that business travelers are the least satisfied segment, suggesting these issues are particularly acute for them. The data provides a clear mandate to investigate and enhance seating comfort, food quality, and the level of customer service provided by staff and crew to mitigate the primary sources of customer complaints.



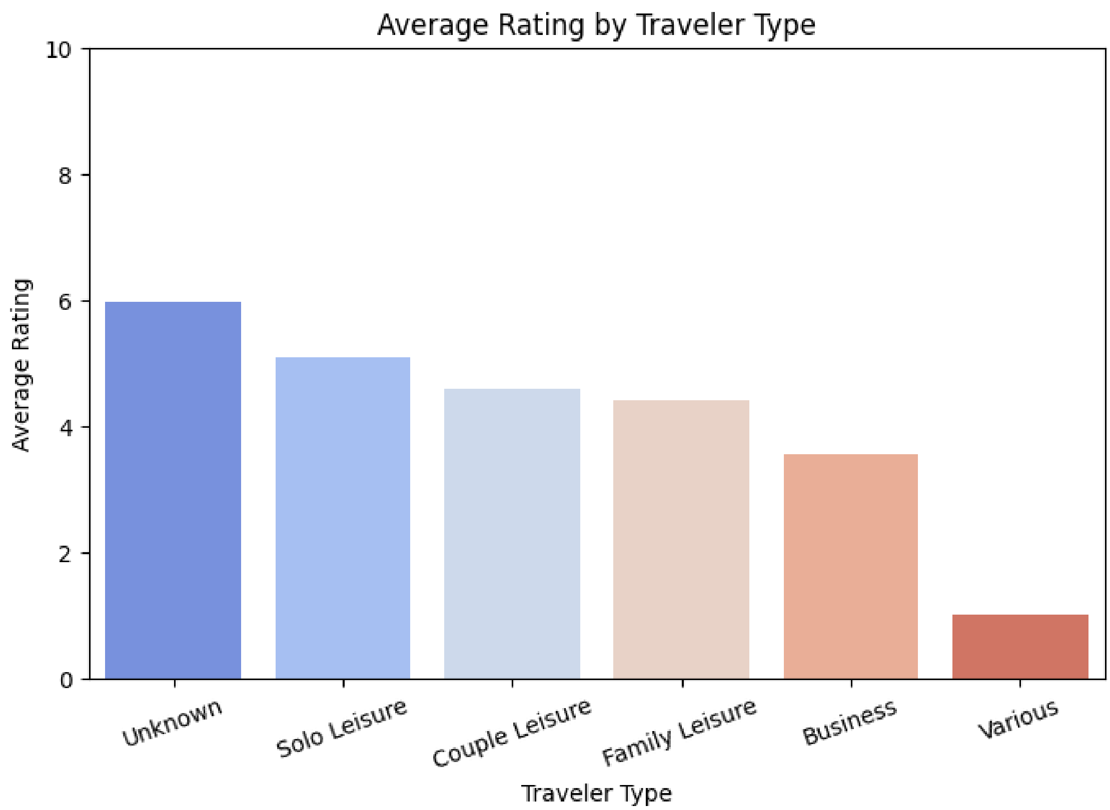
What is the distribution of reviews across different traveler types?

Reasoning for the Question

Analyzing the distribution of traveler types is essential for understanding the airline's primary customer base. By identifying which segments are most vocal, the airline can strategically tailor its services, marketing, and improvement efforts to meet the specific needs of its largest and most engaged customer groups. This insight helps align the airline's business strategy with its actual market perception, whether it's predominantly for leisure, business, or family travel.

Justification and Analysis of the Chart

The bar chart clearly demonstrates that leisure travelers are the dominant source of reviews, with **"Couple Leisure"** and **"Solo Leisure"** being the two most frequent categories. This finding justifies focusing on the vacationer experience, as they form the core of this feedback dataset. While **"Business"** travelers also represent a significant segment, the high volume of **"Unknown"** reviews highlights a potential gap in data collection. The visualization confirms that the airline's feedback loop is driven primarily by those traveling for personal reasons, making their satisfaction a top priority.



How does the average satisfaction rating differ across various traveler types?

Reasoning for the Question

Moving beyond the volume of reviews, this question seeks to identify which specific customer segments are the most and least satisfied with their travel experience. Understanding these satisfaction gaps is critical for the airline to pinpoint specific pain points and opportunities for improvement. By comparing average ratings, the airline can prioritize its resources, focusing on enhancing services for dissatisfied groups (like business travelers) or reinforcing the positive experiences of its happier customers (like leisure travelers) to boost loyalty and brand reputation.

Justification and Analysis of the Chart

This chart clearly reveals a significant disparity in satisfaction across traveler types, justifying a targeted approach to service improvement. The most striking insight is that **"Business" travelers report the lowest average rating** (approximately 3.6 out of 10) among the well-defined segments, indicating a substantial service gap for this valuable customer group. In contrast, leisure travelers ("Solo" and "Couple") exhibit moderate satisfaction with ratings between 4.5 and 5.1. The fact that business travelers are significantly less satisfied than their leisure counterparts provides a clear, actionable directive: the airline must investigate and address the specific factors negatively impacting the corporate travel experience.