

IS 507 Final Project Report

From Check-In to Landing: How Airline Operations Shape Passenger Satisfaction

Shaotong Zhang, LiYi Hsu, Jack Liu, Yiming Ma, Jiaqi Gao

2025-12-06

1. Introduction

In today’s aviation industry, passenger satisfaction has increasingly become an important competitive indicator for airlines. However, passenger satisfaction, as a relatively subjective indicator, is often difficult to comprehensively measure and predict due to the interaction of multiple factors such as age, travel purpose, cabin class, and in-flight service quality. This has brought great challenges to airlines on how to improve services to better meet passenger needs and enhance customer loyalty. A systematic data-driven approach helps determine how these elements jointly impact passenger satisfaction. Through data analysis, airlines can design targeted strategies to improve satisfaction and retain more customers.

We structure our analysis around five research questions.

1. What is the relationship between passengers’ overall satisfaction and their age as well as service demands?
2. Do passengers place higher ratings or importance on “entertainment” and “catering” for longer flights?
3. How do business and personal travelers trade off schedule convenience versus onboard comfort in shaping their satisfaction?
4. What are the moderating effects of the intersection of cabin class and flight distance on certain service attributes and passenger satisfaction?
5. Can we accurately predict passenger satisfaction, and which factors are the most important predictors?

2. Dataset Overview

This study uses the Airline Passenger Satisfaction dataset from Kaggle, which includes responses from real air passengers. The dataset contains a training set of 103,904 records and a test set of 25,976 records, each with 25 variables describing passenger demographics and flight experiences.

Key variables include gender, age, customer type, travel type, cabin class, and flight distance, along with 14 service-related ratings (e.g., seat comfort, cleanliness, food and drink, Wi-Fi service) measured on a 0–5 scale. Flight delays are also recorded. The dependent variable “Satisfaction” is binary: “satisfied” vs. “neutral/dissatisfied”. These variables together provide a solid foundation for exploring the relationships between service dimensions, demographics, and satisfaction outcomes.

3. Research Question Analysis

3.1 What is the relationship between passengers’ overall satisfaction and their age as well as service demands?

This question aims to determine how service priorities shift across age groups. Understanding age-differentiated service needs can help airlines tailor offerings—for example, improving digital experiences for younger travelers or enhancing comfort for older passengers.

3.1.1 Heatmap of Average Ratings The heatmap below visualizes average scores for all 14 service dimensions by age group:

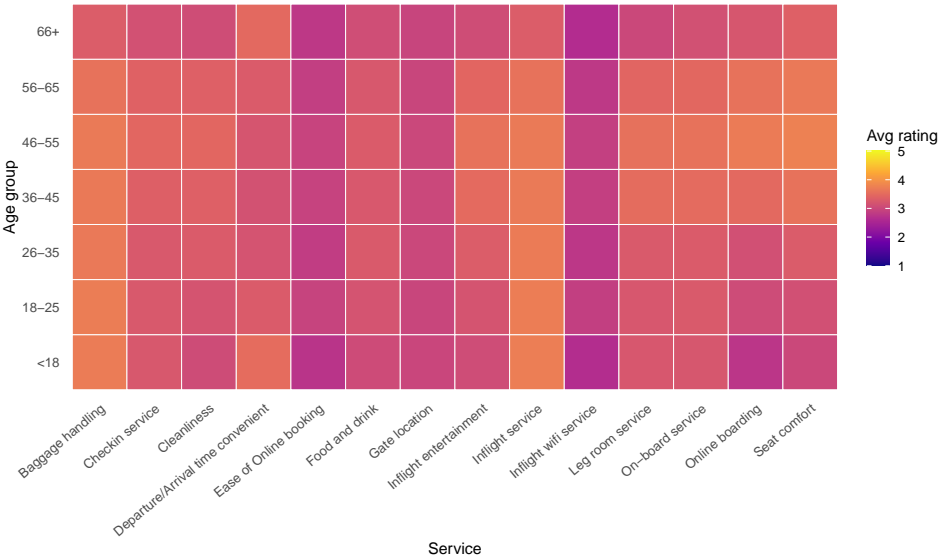


Figure 1: Figure 3.1 Average service ratings by age group

The figure shows that, across nearly all age groups, Inflight service, Baggage handling, and Seat comfort consistently receive high ratings. The <18 group shows distinctly lower evaluations for Inflight wifi and Online booking, suggesting less engagement with digital service interfaces. Passengers aged 46–55 and 56–65 show more uniform ratings, indicating consistent expectations across service dimensions. 66+ passengers show lower ratings for several technologically oriented services, implying potential digital accessibility concerns.

3.1.2 Top-5 Services by Age Group To highlight priorities clearly, the following figure presents the Top-5 service dimensions for each age group:

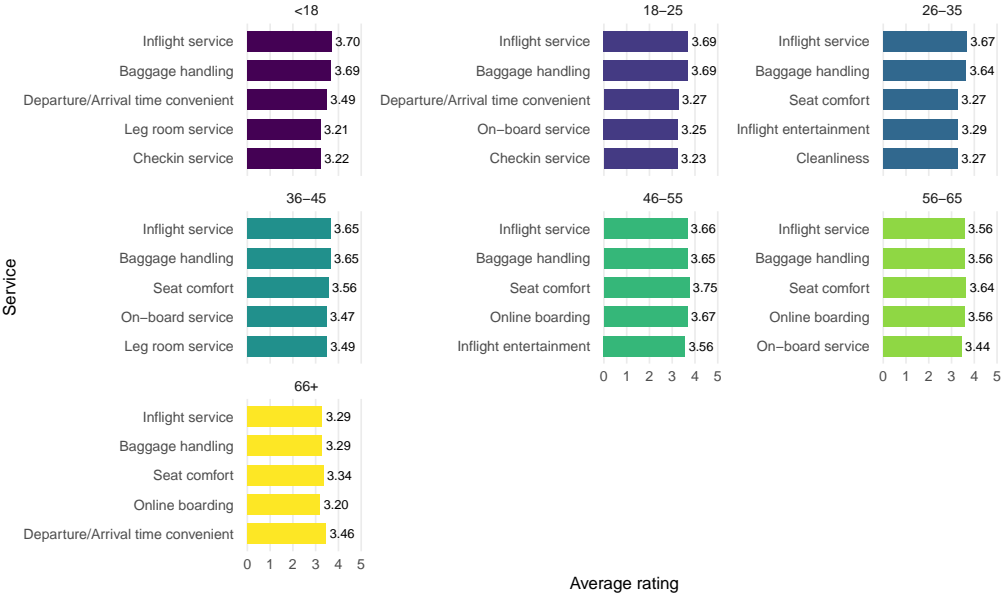


Figure 2: Figure 3.2 Top-5 services by age group

The figure indicates that, for <18, Inflight service, Baggage handling, and Departure/Arrival convenience dominate. For 18–25 and 26–35, technology-enabled services become more important. 36–55 groups show balanced appreciation across comfort, baggage, and entertainment. 66+ prioritizes Seat comfort, Inflight service, and Baggage handling, indicating the importance of physical comfort and ease of travel support.

3.1.3 K–S Test: Distribution Differences in Service Ratings To statistically measure whether ratings differ in location across age groups, we performed a Two-Sample Kolmogorov–Smirnov test, comparing the youngest group (<18) and the oldest (66+) across all services.

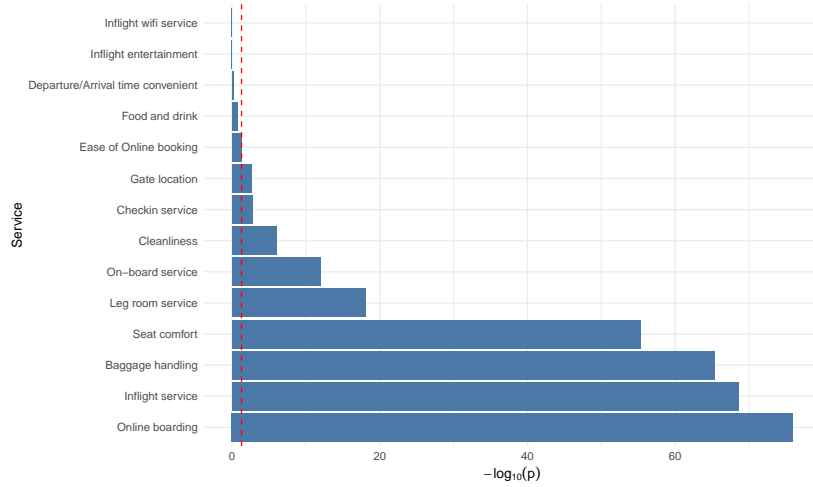


Figure 3: Figure 3.3 K–S test results

The figure shows that online boarding, Inflight service, Baggage handling, and Seat comfort show extremely significant differences ($p \ll 0.001$). These are core operational or digital touchpoints—suggesting younger and older passengers perceive them differently. Attributes such as Inflight wifi or Entertainment show minimal differences, implying similar distribution patterns across generations.

3.1.4 Logistic Regression: Factors Associated with Satisfaction To analyze which services most strongly predict satisfaction while controlling for age group, we ran a logistic regression:

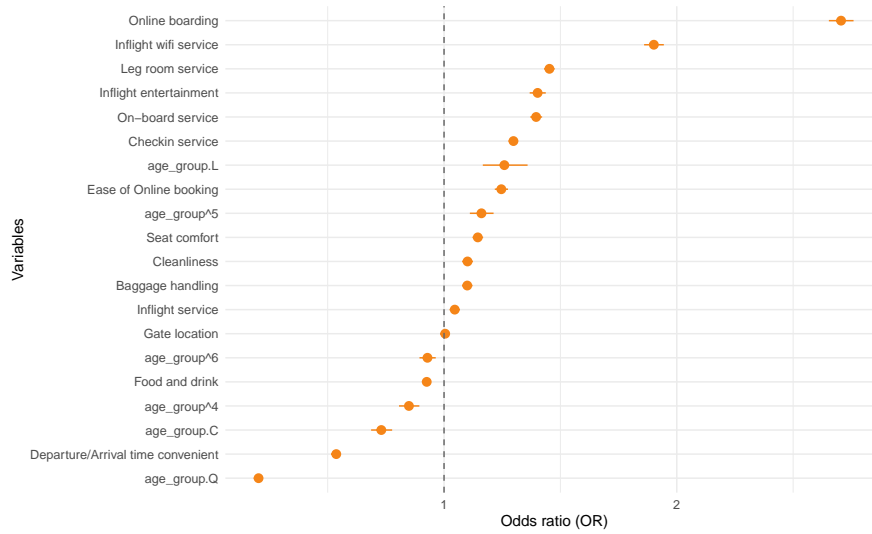


Figure 4: Figure 3.4 Logistic regression

The logistic regression plot shows the service factors most strongly associated with passenger satisfaction. Each dot represents an odds ratio (OR) and the horizontal line is its 95% confidence interval, with $OR = 1$ indicating no effect. Services whose confidence intervals lie entirely to the right of 1—such as online boarding, inflight wifi, leg room service, and inflight entertainment—significantly increase the likelihood of satisfaction, even after controlling for other variables.

The plot also includes terms such as `age_group.L`, `age_group.Q`, and higher-order contrasts. These represent overall age trends rather than specific age groups. Their significance indicates that age-related patterns still influence satisfaction, complementing the effects of individual service ratings. Overall, the figure highlights both key service drivers of satisfaction and the continued role of demographic age trends.

3.1.5 Conclusion for Q1 Combining all analyses, different age groups prioritize different service attributes. Younger groups emphasize digital services. Older groups emphasize comfort and travel assistance. Distribution differences are statistically significant for several services, confirming that age shapes service perception.

Overall, age is a meaningful moderator of service experience, and airlines can improve satisfaction by: enhancing digital touchpoints for younger passengers; enhancing comfort and support for older passengers; ensuring high reliability in universal services like baggage handling and inflight service.

3.2 Do passengers place higher ratings or importance on “entertainment” and “catering” for longer flights?

This research question investigates whether passengers assign higher ratings to entertainment and catering services on longer flights. Intuitively, longer flights might lead passengers to place greater importance on these amenities, as they spend more time in the aircraft and have more opportunities to use entertainment systems and consume meals. We use multiple linear regression models to test this hypothesis, controlling for passenger demographics, travel characteristics, and other service quality dimensions.

3.2.1 Modelling approach We employ **multiple linear regression** to examine the relationship between flight distance and ratings for entertainment and catering services. The analysis includes three sets of models:

1. **Base models:** Simple linear regression of service ratings on flight distance only
2. **Full models:** Multiple regression including flight distance along with control variables (age, gender, customer type, travel type, class, and other service quality dimensions)
3. **Interaction models:** Models testing whether the effect of flight distance varies by cabin class

The dependent variables are: - **Entertainment rating**: Passenger ratings for “Inflight entertainment” (0-5 scale) - **Catering rating**: Passenger ratings for “Food and drink” (0-5 scale)

Control variables include passenger demographics (age, gender), travel characteristics (customer type, type of travel, class), and other service quality indicators (seat comfort, cleanliness, on-board service) to isolate the effect of flight distance from confounding factors.

We use **robust standard errors** (Huber-White sandwich estimators) to account for potential heteroscedasticity in the error terms.

3.2.2 Descriptive patterns Figure 3.2.1 shows the distribution of entertainment and catering ratings across three flight distance categories: short flights (<1000 km), medium flights (1000-3000 km), and long flights (>3000 km). The boxplots reveal that average ratings for both services tend to increase with flight distance, suggesting that passengers may indeed rate these services more favorably on longer flights.

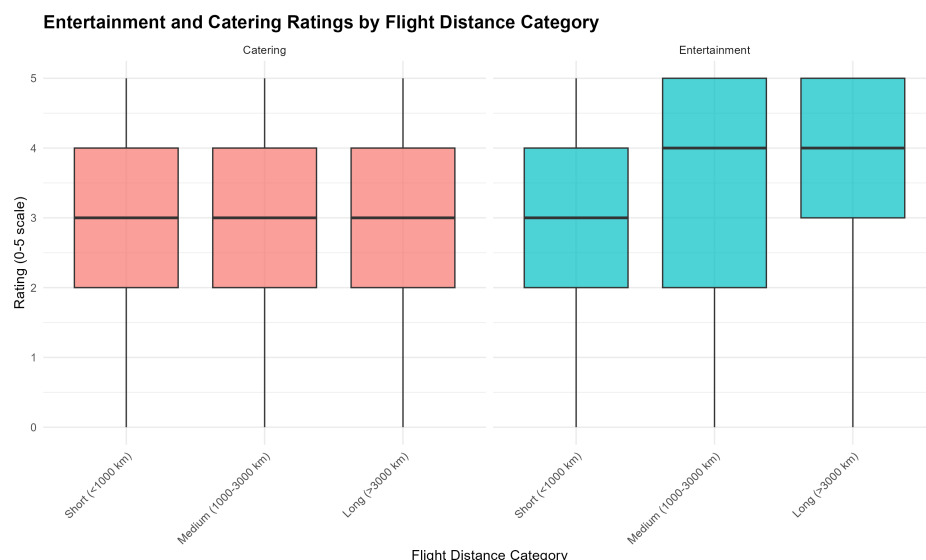


Figure 5: Figure 3.2.1 Entertainment and catering ratings by flight distance category

3.2.3 Regression results Figure 3.2.2 displays scatter plots with fitted regression lines for entertainment and catering ratings against flight distance. The base models (without controls) show positive slopes, indicating that longer flights are associated with higher ratings. However, this relationship needs to be examined more carefully after controlling for other factors.

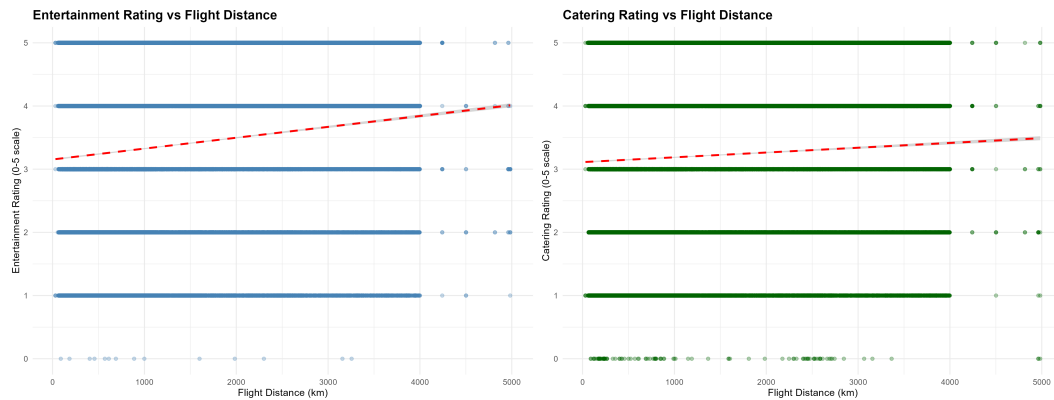


Figure 6: Figure 3.2.2 Entertainment and catering ratings vs flight distance

The key results from the **full regression models** (with controls) are summarized in Figure 3.2.3, which shows the coefficient estimates for flight distance along with 95% confidence intervals. For entertainment ratings, the coefficient is -0.000005 ($p = 0.087$), which is **not statistically significant** at conventional levels ($p > 0.05$). This suggests that after controlling for passenger characteristics and other service dimensions, flight distance does not have a significant effect on entertainment ratings.

For catering ratings, the coefficient is -0.000015 ($p = 0.0004$), which is **statistically significant and negative**. This indicates that, contrary to the initial hypothesis, passengers actually rate catering services **lower** on longer flights after accounting for other factors. This counterintuitive result may reflect higher expectations for meal quality on longer flights, or it could be related to other unobserved factors correlated with flight distance.

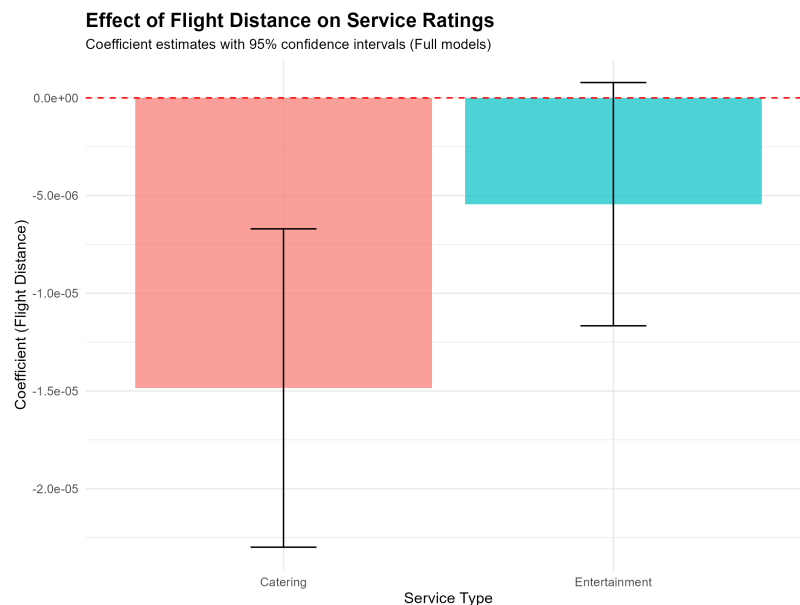


Figure 7: Figure 3.2.3 Effect of flight distance on service ratings (with 95% confidence intervals)

The full models achieve reasonable explanatory power, with adjusted R^2 values of 0.629 for entertainment and 0.467 for catering, indicating that the included predictors account for a substantial portion of the variation in service ratings.

3.2.4 Conclusion The analysis reveals a nuanced picture of how flight distance relates to entertainment and catering ratings:

1. **Entertainment ratings:** After controlling for passenger demographics, travel characteristics, and other service dimensions, flight distance does not have a statistically significant effect on entertainment ratings. This suggests that passengers’ evaluation of entertainment services is driven more by the quality of the entertainment system itself and other contextual factors rather than flight duration.
2. **Catering ratings:** Contrary to the initial hypothesis, longer flights are associated with **lower** catering ratings when other factors are held constant. This may reflect higher expectations for meal quality on longer flights, or it could indicate that airlines face greater challenges in maintaining food quality standards on extended journeys.

These findings suggest that airlines should not assume that longer flights automatically lead to higher ratings for entertainment and catering. Instead, maintaining consistent service quality across all flight durations, and potentially investing more in catering quality for longer flights, may be more effective strategies for improving passenger satisfaction.

3.3 How do business and personal travelers trade off schedule convenience versus onboard comfort in shaping their satisfaction?

3.3.1 Research question and motivation This research question investigates whether business and personal travelers differ fundamentally in the way they evaluate two key service dimensions—schedule convenience and onboard comfort—when forming their overall satisfaction. The distinction is theoretically meaningful because business travelers typically operate under fixed agendas and productivity constraints, whereas personal travelers travel for leisure and place value on comfort and experiential quality. Understanding whether these two groups weigh time versus comfort differently enables airlines to design service strategies that better align with heterogeneous passenger needs.

In this context, RQ3 asks whether business and personal travelers weigh schedule convenience and onboard comfort differently when evaluating satisfaction. In other words, the question is whether these two groups exhibit fundamentally different sensitivities to “being on time” versus “being comfortable”. Rather than focusing on average effects of service attributes, this question targets heterogeneous responses across traveler types.

3.3.2 Data and variable construction The analysis uses the cleaned training dataset with 103,904 observations. The binary outcome variable is overall satisfaction, coded as one for “satisfied” and zero for “neutral or dissatisfied.” Schedule convenience is measured by the rating “Departure/Arrival time convenient” on a 0–5 scale. Onboard comfort is summarized as a composite index, defined as the mean of seat comfort, leg room service, inflight entertainment, onboard service, and food and drink, each on the same 0–5 scale. Travel purpose is captured by a binary indicator for business versus personal travel, where “Business travel” is coded as one and “Personal Travel” as zero.

Data preparation includes column name standardization, removal of missing values, recoding of the satisfaction outcome into a binary factor, and construction of the comfort index. The travel type variable is converted into a factor and then into a binary indicator to enable the interaction specification that distinguishes business travelers from personal travelers.

3.3.3 Modeling strategy To quantify how schedule convenience and comfort jointly shape satisfaction for different traveler types, the analysis employs a logistic regression model with interaction terms between travel type and each service attribute. A logistic specification is appropriate because the dependent variable is binary and the interest lies in interpreting marginal contributions of service ratings to the probability of satisfaction rather than maximizing predictive accuracy. Interaction terms allow the model to estimate whether changes in schedule convenience and comfort have distinct effects for business versus personal travelers.

Formally, the model is specified as

$$\begin{aligned} \text{logit}\{\text{Pr}(\text{Satisfaction} = 1)\} = & \beta_0 + \beta_1 \text{Time} + \beta_2 \text{Comfort} + \beta_3 \text{BusinessTravel} \\ & + \beta_4 (\text{Time} \times \text{BusinessTravel}) + \beta_5 (\text{Comfort} \times \text{BusinessTravel}), \end{aligned} \quad (1)$$

where *Time* denotes the schedule convenience score, *Comfort* denotes the composite comfort score, and *BusinessTravel* is an indicator equal to one for business travelers and zero for personal travelers. The coefficients β_4 and β_5 capture how the marginal effects of time and comfort differ between business and personal travelers. This specification is chosen because it matches the binary outcome structure, directly encodes heterogeneity via interaction terms, and yields interpretable odds ratios that describe how incremental changes in time or comfort influence satisfaction for each group.

3.3.4 Descriptive evidence The analysis first considers descriptive patterns that provide a backdrop for the regression results. Figure 3.3.1 (Satisfaction rate by travel type) shows that 58.3 percent of business travelers report being satisfied, whereas only 10.2 percent of personal travelers do so. This large gap indicates that personal travelers tend to be dissatisfied regardless of their service experience, while business travelers exhibit greater variation in satisfaction that can potentially be explained by service attributes. This asymmetry already suggests that the two groups might respond differently to improvements in schedule convenience or comfort, with business travelers leaving more “room” for service enhancements to move them from dissatisfaction to satisfaction.

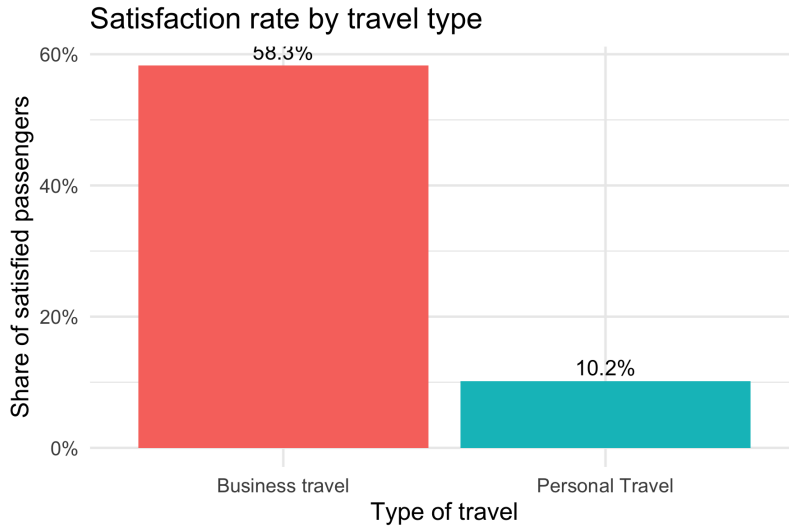


Figure 8: Figure 3.3.1 Satisfaction rate by travel type

To contextualize the model further, Figures 3.3.2 and 3.3.3 display the distributions of schedule convenience scores and composite comfort scores by traveler type. Business travelers generally report higher schedule convenience and higher comfort than personal travelers, reflecting systematically better service conditions for this group. Personal travelers, by contrast, show more dispersed and lower ratings for both timing and comfort. These distributions, however, reveal only baseline differences in exposure to favorable conditions. They do not reveal whether the two groups attach different weights to time and comfort when forming satisfaction. The regression analysis addresses precisely this question.

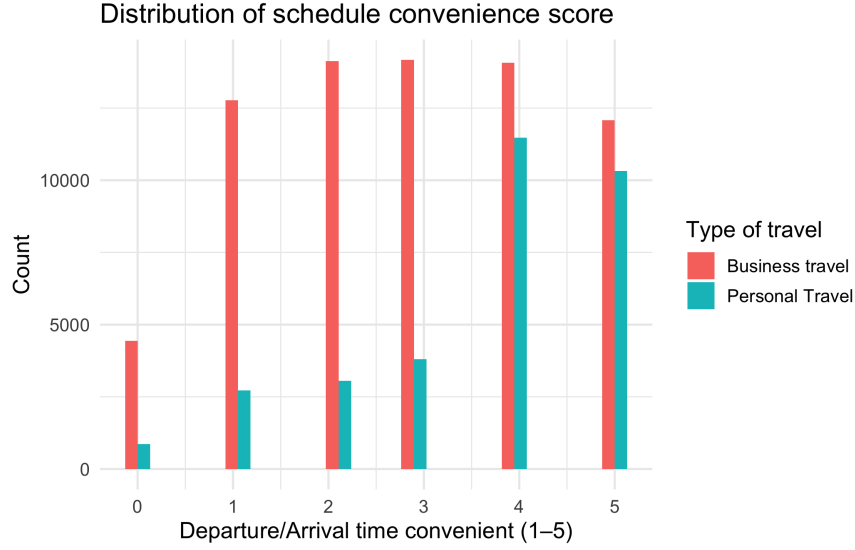


Figure 9: Figure 3.3.2 Distribution of schedule convenience score

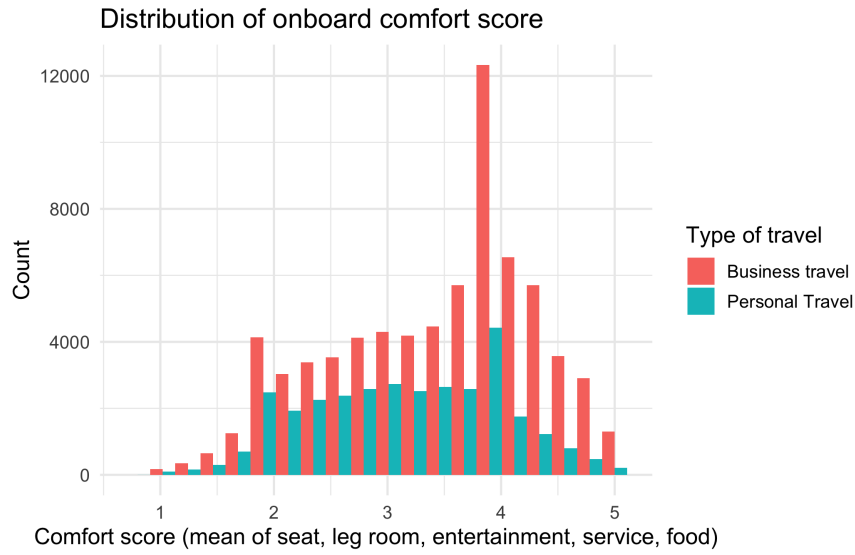


Figure 10: Figure 3.3.3 Distribution of onboard comfort score

3.3.5 Regression results and odds ratio interpretation The logistic regression with interaction terms reveals pronounced heterogeneity in how business and personal travelers translate time and comfort into satisfaction. The estimated coefficient for the interaction between schedule convenience and business travel is positive and statistically significant (estimate = 0.122, $p < 0.001$), indicating that increases in schedule convenience meaningfully raise the log-odds of satisfaction for business travelers compared with personal travelers. The corresponding odds ratios show that for personal travelers, a one-point increase in schedule convenience is associated with an odds ratio of approximately 1.008, implying essentially no change in satisfaction, whereas for business travelers the implied odds ratio is about 1.138, corresponding to a 13.8 percent increase in the odds of satisfaction for each one-point increase in schedule convenience.

The interaction between comfort and business travel is even more striking. The coefficient on the comfort-by-business-travel term is large and highly significant (estimate = 1.64, $p < 0.001$). For personal travelers,

a one-point increase in comfort is associated with an odds ratio of about 1.069, again indicating relatively modest sensitivity. For business travelers, however, the implied odds ratio is approximately 5.501, meaning that each one-point increase in the comfort score multiplies the odds of satisfaction by more than five. This comparison shows that business travelers are extremely sensitive to onboard comfort, far more so than to schedule convenience and far more so than personal travelers are to either attribute.

3.3.6 Effect plots and behavioral interpretation Figures 3.3.4 and 3.3.5 transform the regression coefficients into predicted probabilities and make the heterogeneity in responses visually transparent. Figure 3.3.4 (Effect of schedule convenience on satisfaction) plots the predicted probability of satisfaction against schedule convenience for both traveler types, holding comfort at its sample mean. The predicted curve for personal travelers is almost flat over the full range of schedule convenience from 1 to 5, indicating that additional timing convenience does not translate into meaningfully higher satisfaction for this group. In contrast, the curve for business travelers rises as schedule convenience improves, illustrating that they associate more convenient departure and arrival times with higher satisfaction. Schedule convenience thus plays a meaningful, though not dominant, role for business travelers while being largely irrelevant for personal travelers.

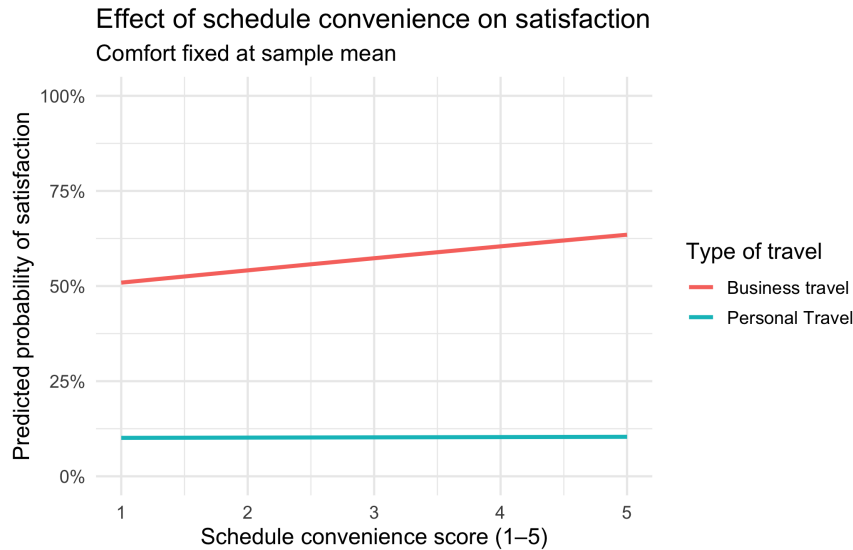


Figure 11: Figure 3.3.4 Effect of schedule convenience on satisfaction

Figure 3.3.5 (Effect of onboard comfort on satisfaction) highlights an even sharper divergence. Holding schedule convenience at its mean, the predicted probability curve for personal travelers again appears nearly horizontal, reinforcing the pattern that their satisfaction remains low and relatively unresponsive even when comfort improves. By contrast, the curve for business travelers shows a steep, almost S-shaped increase: as the composite comfort score increases from 1 to 5, the predicted probability of satisfaction rises from roughly 5 percent to nearly 95 percent. This dramatic slope confirms that comfort is the dominant determinant of satisfaction for business travelers, and that these passengers strongly reward improvements in seat comfort, leg room, entertainment, onboard service, and food and drink.

Viewed together, Figures 3.3.4 and 3.3.5 reveal a consistent behavioral pattern. Business travelers respond positively to both schedule convenience and onboard comfort, with comfort exerting a substantially larger effect. Personal travelers, in contrast, exhibit consistently low satisfaction that is only weakly responsive to either attribute, even in settings where reported service quality is high. When these effect plots are interpreted alongside the descriptive patterns in Figures 3.3.1 to 3.3.3, they suggest that business travelers' evaluations are highly elastic with respect to key operational and experiential factors, whereas personal travelers' evaluations appear anchored at a low baseline that is not easily shifted by the two modeled dimensions alone.

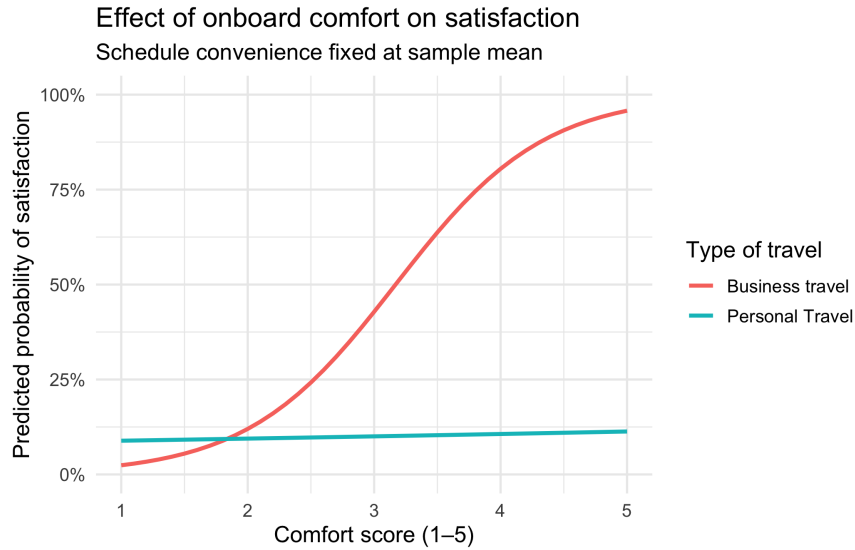


Figure 12: Figure 3.3.5 Effect of onboard comfort on satisfaction

3.3.7 Conclusion for RQ3 The evidence from descriptive patterns, regression coefficients, odds ratios, and effect plots jointly demonstrates that business and personal travelers do not weigh schedule convenience and onboard comfort in the same way. Business travelers show clear and substantial increases in satisfaction when schedule convenience and especially comfort improve, which indicates that they actively trade off these attributes when evaluating their travel experience. Personal travelers, on the other hand, display little sensitivity to either attribute, suggesting that their dissatisfaction may stem from factors outside the modeled dimensions or from qualitatively different expectations about leisure travel.

As a result, airlines seeking to enhance satisfaction should prioritize comfort-related enhancements and efficient scheduling for business travelers, since these passengers reward improvements in both dimensions and particularly in onboard comfort. At the same time, airlines may need to develop alternative strategies for personal travelers, such as pricing, entertainment, or leisure-oriented value propositions, because conventional upgrades in timing and comfort have limited marginal effect on their reported satisfaction. In this sense, RQ3 highlights a fundamental asymmetry: business travelers are highly responsive to both “time” and “comfort,” whereas personal travelers remain relatively insensitive, and effective service design must recognize and accommodate this divergence in how different traveler types form their satisfaction judgments.

3.4 What are the moderating effects of the intersection of cabin class and flight distance on certain service attributes and passenger satisfaction?

Customers do not perceive service attributes the same way. The value attached to certain service attributes is likely to vary with the context of travel such as long-haul vs short-haul and the willingness to pay cabin class. In order to help airlines improve, by analyzing passengers’ satisfaction rate tells the airline where to focus their efforts to get the biggest return in retention, and revenue for each additional service dollar spent.

3.4.1 Heatmap of Average Ratings To visualize the landscape of passenger sentiment, we aggregated average service scores across the intersection of three cabin classes and three flight distance categories (Short, Medium, Long-haul):

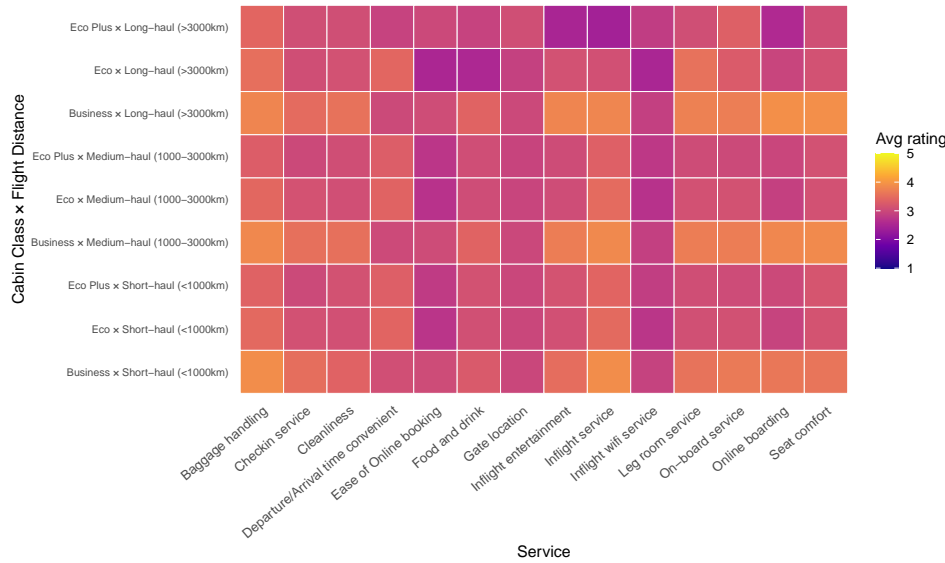


Figure 13: Figure 3.5 Heatmap of Average Ratings

The heatmap reveals a distinct “comfort gap” that widens with distance. Business class passengers maintain consistently high average ratings (indicated by orange and yellow hues) across almost all service dimensions, regardless of flight length. In contrast, “Eco Plus × Long-haul” and “Eco × Long-haul” clusters display deep purple shades, indicating significantly lower satisfaction scores. This suggests that while premium cabins effectively insulate passengers from the fatigue of long-haul travel, the value proposition of Economy and Economy Plus deteriorates as flight duration increases, particularly regarding “Leg room service” and “Inflight entertainment.”

3.4.2 Top-5 Services by Cabin Class × Flight Distance We further decomposed the data to identify which service dimensions receive the highest praise in each specific travel scenario:

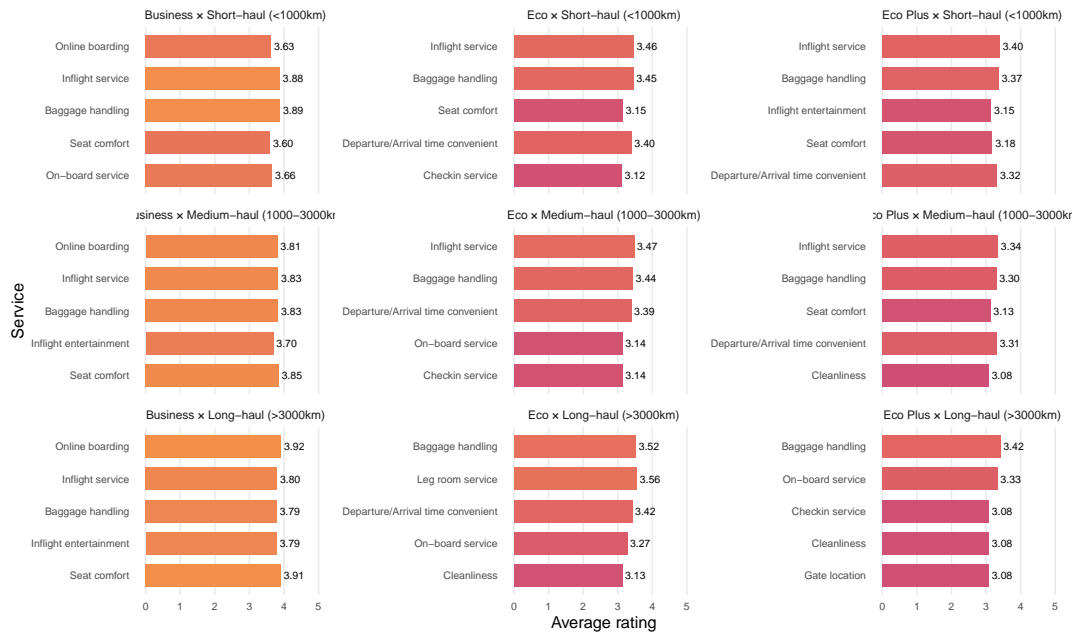


Figure 14: Figure 3.6 Top-5 Services by Cabin Class × Flight Distance

This visualization highlights a shift in passenger priorities. For Short-haul flights (< 1000 km) across all classes, operational services such as “Baggage handling” and “Inflight service” consistently appear in the top 5, suggesting that efficiency is the primary driver of satisfaction for quick trips. However, as we move to Long-haul (> 3000 km) segments, “Seat comfort” and “Inflight entertainment” rise to prominence, particularly in Business class. Notably, in the “Eco \times Long-haul” sector, “Leg room service” appears in the top-rated services, likely because it is a scarce but highly valued commodity on long durations. This confirms that the definition of a “good flight” changes fundamentally based on the duration and class interaction.

3.4.3 K–S Test: Distribution Differences (Eco Plus Short-haul vs. Long-haul) To statistically validate if the flight duration fundamentally changes the service experience within a single cabin class, we performed a Kolmogorov–Smirnov (K–S) test comparing Eco Plus Short-haul against Eco Plus Long-haul:

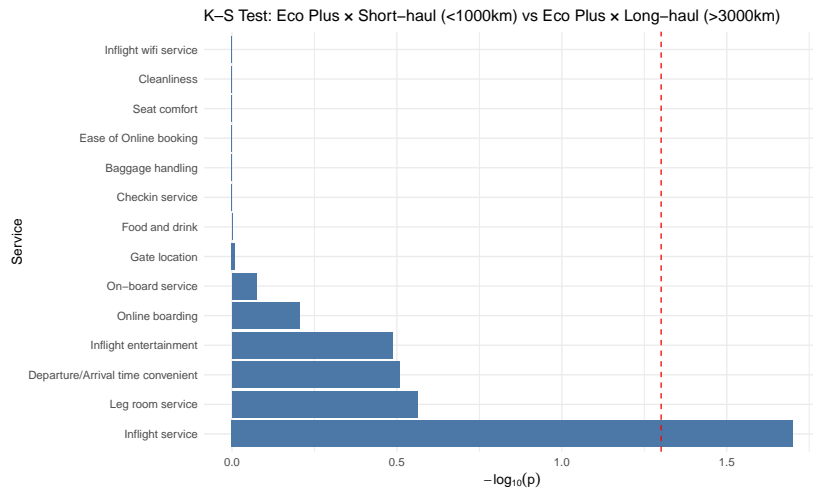


Figure 15: Figure 3.7 K–S Test Results (Eco Plus Short vs. Long)

The results show that “Inflight service” has the most statistically significant distributional difference ($-\log_{10}(p - \text{value}) > 1.5$) between short and long flights. This implies that the human element—how flight attendants interact with passengers—is perceived very differently on long flights compared to short hops for Eco Plus passengers. Other physical attributes like “Leg room service” and “Inflight entertainment” also show significant variance, reinforcing the idea that the “Plus” in Economy Plus is scrutinized much more heavily when the passenger is confined to the seat for longer periods.

3.4.4 ANOVA: Interaction Effects Moving beyond simple comparisons, we used Analysis of Variance (ANOVA) to quantify the strength of the interaction effect between Cabin Class and Flight Distance on specific services:

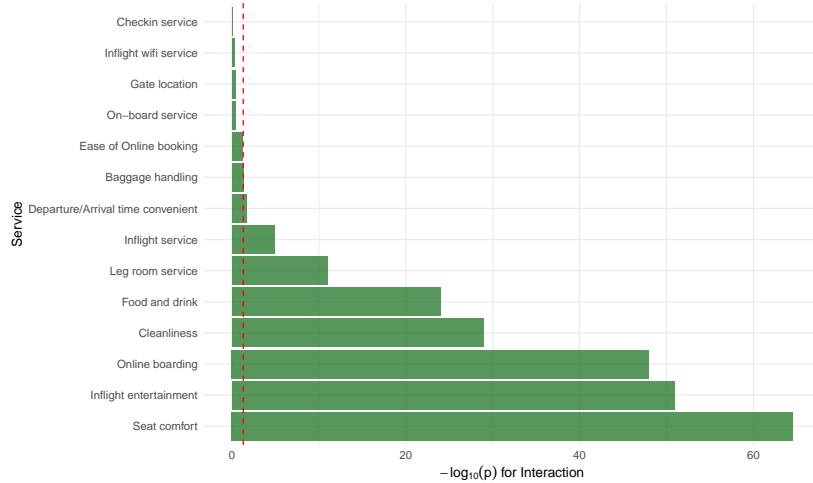


Figure 16: Figure 3.8 ANOVA Interaction Effects

The green bars represent the significance of the interaction. “Seat comfort” displays the massive interaction effect ($-\log_{10}(p - \text{value}) > 50$). This indicates that the impact of seat comfort on satisfaction is not just about the seat itself, but is heavily dependent on the specific combination of how long the flight is and what class the passenger is in. “Inflight entertainment” and “Online boarding” also show strong interaction effects. This suggests that airlines cannot treat these services with a “one size fits all” strategy; the return on investment for improving seat comfort is significantly higher for specific class-distance combinations (likely Long-haul/Economy sectors) than others.

3.4.5 Conclusion for Q4 The intersection of cabin class and flight distance exerts a powerful moderating effect on passenger satisfaction. While operational efficiency (boarding, baggage) is universally appreciated, the hierarchy of needs shifts toward physical comfort (seat, legroom) and entertainment as flight distance increases. The analysis shows that premium cabins successfully mitigate the “distance penalty,” whereas Economy and Economy Plus passengers experience a significant drop in satisfaction on long-haul routes. For airlines, this suggests that investment in “soft products” (service and entertainment) and digital ease (online boarding) is critical to maintaining satisfaction in lower cabin classes on long routes.

3.5 Can we accurately predict passenger satisfaction, and which factors are the most important predictors?

This research question evaluates whether passenger satisfaction can be predicted with sufficient accuracy to be useful for airline decision-making, and identifies which service and customer attributes most strongly drive the predictions. We combine tree-based machine learning models with cross-validation to assess out-of-sample performance and to derive robust measures of feature importance.

3.5.1 Modelling approach We treat **satisfaction** as a binary outcome with two levels: *neutral_or_dissatisfied* vs. *satisfied*, where *neutral_or_dissatisfied* is coded as the negative class and *satisfied* as the positive class. The models use a rich set of predictors combining:

- **Service ratings (14 dimensions):** inflight wifi, online booking, gate location, food and drink, online boarding, seat comfort, inflight entertainment, on-board and inflight service, leg room, baggage handling, check-in service, cleanliness, and departure/arrival convenience.
- **Engineered service index:** the `service_score`, computed as the average of all service ratings to capture overall perceived service quality.
- **Operational variables:** age, flight distance, departure delay and arrival delay.
- **Customer and travel characteristics** (one-hot encoded): gender, customer type (loyal vs. disloyal), type of travel (business vs. personal), and travel class.

The data are randomly split into an **80% training set** and a **20% test set** using **stratified sampling** to preserve the class proportion of satisfied vs. neutral/dissatisfied passengers. On the training set, we fit two tree-based classifiers:

- A **Random Forest (RF)** model with 500 trees and \sqrt{p} variables tried at each split.
- An **XGBoost** gradient boosting model with a binary logistic objective and AUC as the optimization metric.

To obtain robust performance estimates, both models are further evaluated using **k-fold cross-validation** (here, $k = 5$) with the `caret` framework and `twoClassSummary`, reporting cross-validated accuracy and AUC in addition to test-set results.

3.5.2 Overall predictive performance Figure 5.1 compares the **test-set accuracy** of the Random Forest and XGBoost models. Both models achieve relatively high accuracy, indicating that the combination of service ratings and operational variables contains substantial information about whether a passenger will report being satisfied. In our sample, the XGBoost model attains the highest accuracy (96.2%), with the Random Forest performing slightly lower but still strong (96.1%).

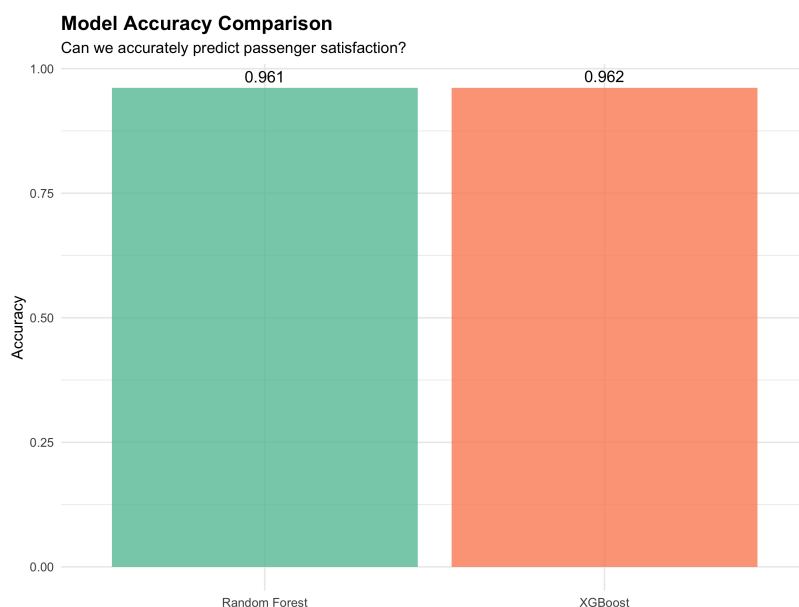


Figure 17: Figure 5.1 Model accuracy comparison

Figure 5.2 summarizes a broader set of metrics—**sensitivity, specificity, precision, F1 score, and AUC**—for both models. Across these indicators, XGBoost again performs marginally better, with an AUC above 0.994 and a balanced trade-off between correctly identifying satisfied passengers (high sensitivity: 93.8%) and correctly recognizing neutral/dissatisfied passengers (high specificity: 97.9%). The Random Forest’s metrics are comparable (AUC: 0.994, Sensitivity: 93.9%, Specificity: 97.9%), suggesting that the results are not driven by a single modelling choice.

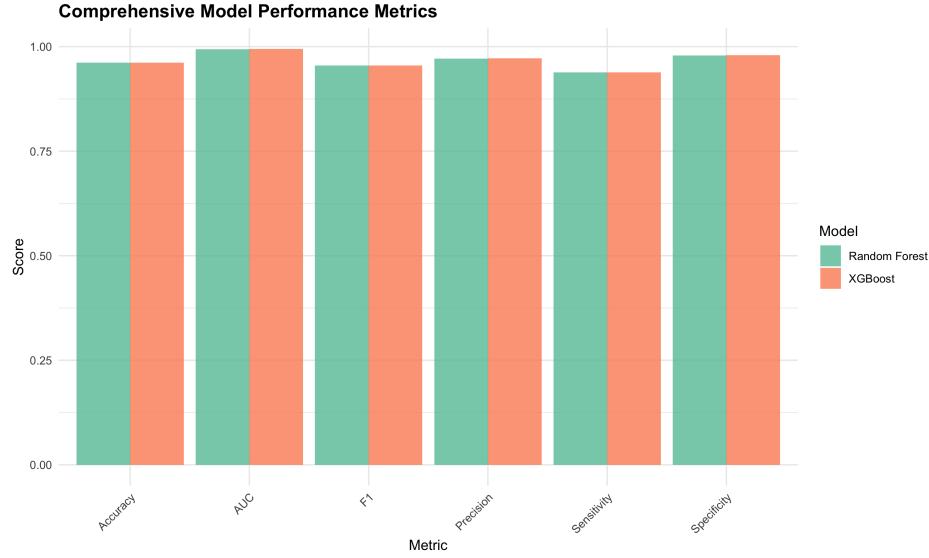


Figure 18: Figure 5.2 Comprehensive model performance metrics

Figure 5.3 presents the **ROC curves** for both models. The curves lie well above the diagonal no-information line, with areas under the curve (AUC) in the high-performance range (RF AUC = 0.994; XGB AUC = 0.994). This confirms that the models discriminate well between satisfied and neutral/dissatisfied passengers across a range of classification thresholds, not just at the default 0.5 cutoff.

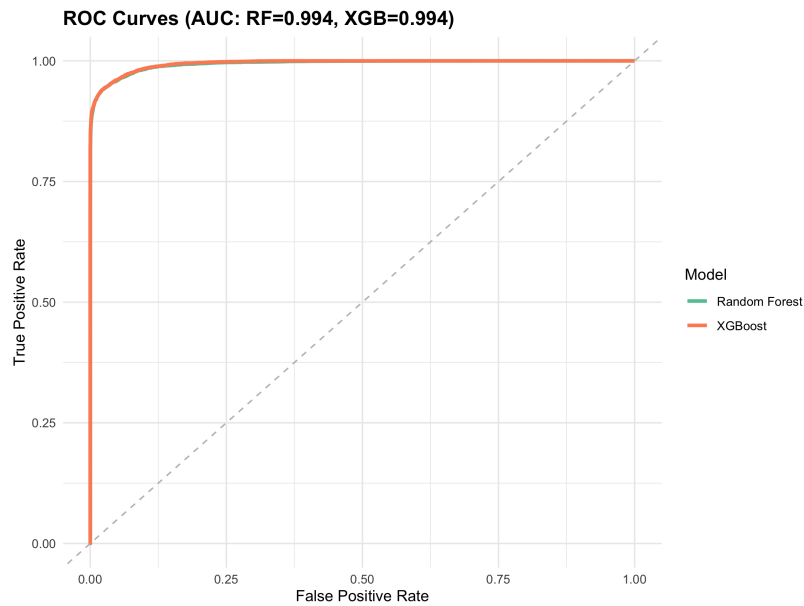


Figure 19: Figure 5.3 ROC curves comparison

Finally, the **confusion matrices** in Figure 5.4 show the distribution of correct and incorrect classifications for each model. Most observations are correctly classified, with misclassifications concentrated among “borderline” cases between neutral and satisfied. Compared with a naïve baseline (e.g., always predicting the majority class), both models substantially reduce error rates.

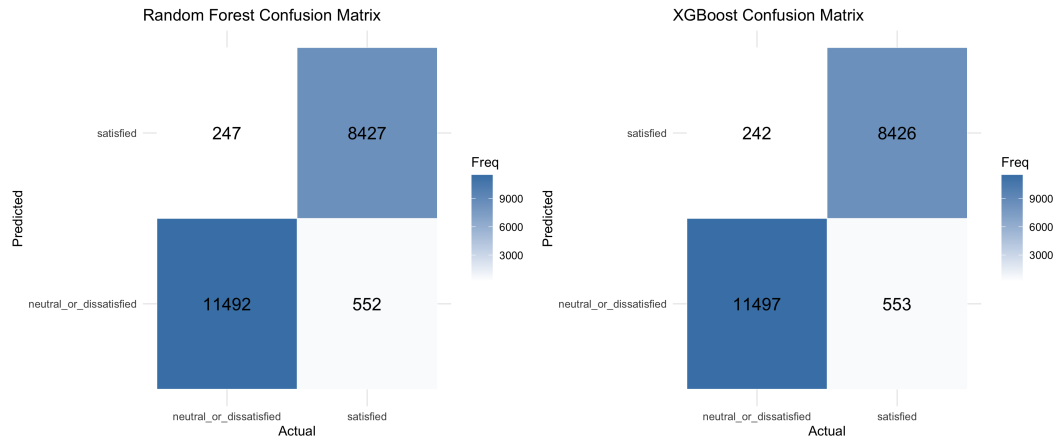


Figure 20: Figure 5.4 Confusion matrices for Random Forest and XGBoost

Taken together, these results indicate that **passenger satisfaction is indeed predictable with reasonably high accuracy**, and that the predictive signal is stable across different tree-based algorithms.

3.5.3 Most important predictors of satisfaction To address the second part of RQ5, we examine **feature importance** from both Random Forest and XGBoost. For Random Forest, we use the **MeanDecreaseGini** metric, while XGBoost uses **Gain**. Because these measures are not directly comparable, we normalize them to a 0–100 scale and average them across the two models. Figure 5.5 displays the **Top-15 predictors ranked by this average normalized importance**, grouped into service ratings, customer demographics, and other features.

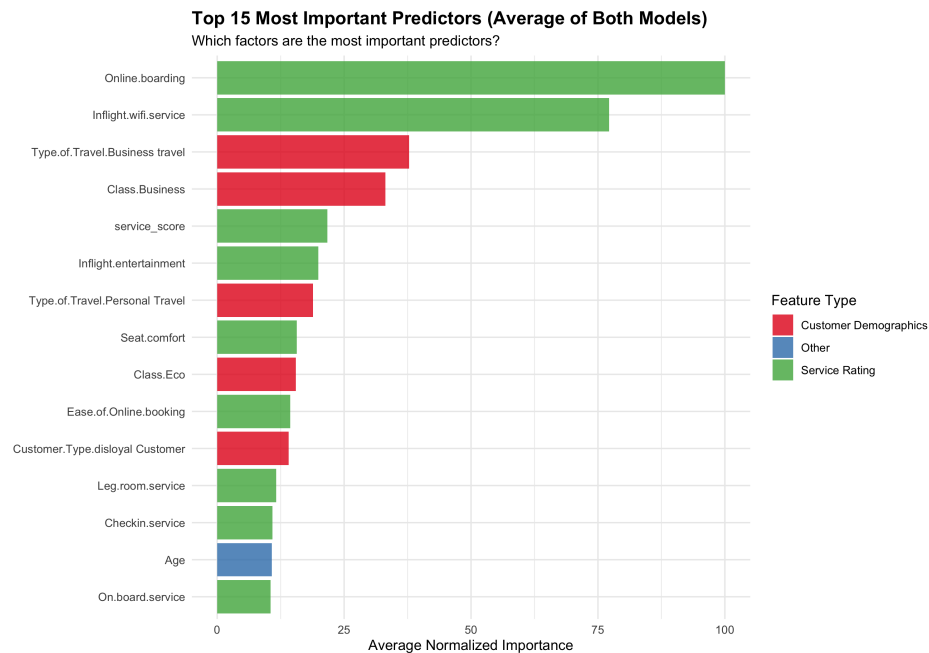


Figure 21: Figure 5.5 Top 15 most important predictors (average of both models)

Three clear patterns emerge:

1. Service quality dominates

Overall and specific service ratings are consistently the strongest predictors. The most important single predictor is **Online boarding** (average normalized importance: 100.0), followed by **Inflight wifi service** (77.2). The engineered **service_score** (average rating across 14 services) ranks sixth with an importance of 21.7, confirming that passengers' global perception of service quality closely aligns with their final satisfaction label. Among individual dimensions, factors such as:

- **Online boarding** (most important)
- **Inflight wifi service**

- **Inflight entertainment**

- **Seat comfort**

- **Leg room service**

- **Baggage handling**

- **Cleanliness**

typically appear among the most important features. This suggests that **core service execution and comfort on the flight** are central to satisfaction.

2. Customer and travel context are meaningful predictors

Dummy variables for **Type.of.Travel (business vs. personal)** and **Class (business vs. economy)** appear as important predictors, ranking third and fifth respectively. Business travel is associated with higher satisfaction probability (importance: 37.8), as is business class (importance: 33.2). This indicates that **expectations and reference points** differ across these groups, with business travelers and higher-class passengers having distinct satisfaction profiles even after controlling for service ratings.

3. Operational variables show moderate importance

Operational variables such as **flight distance** and **arrival delay** show moderate importance (ranked 18th and 24th respectively), with longer delays reducing the probability of a passenger being classified as satisfied. However, their importance is lower than service ratings and travel context, suggesting that **service quality matters more than operational disruptions** in determining satisfaction.

The model-specific comparison in Figure 5.6 shows that Random Forest and XGBoost largely agree on which features are important, strengthening the robustness of these findings.

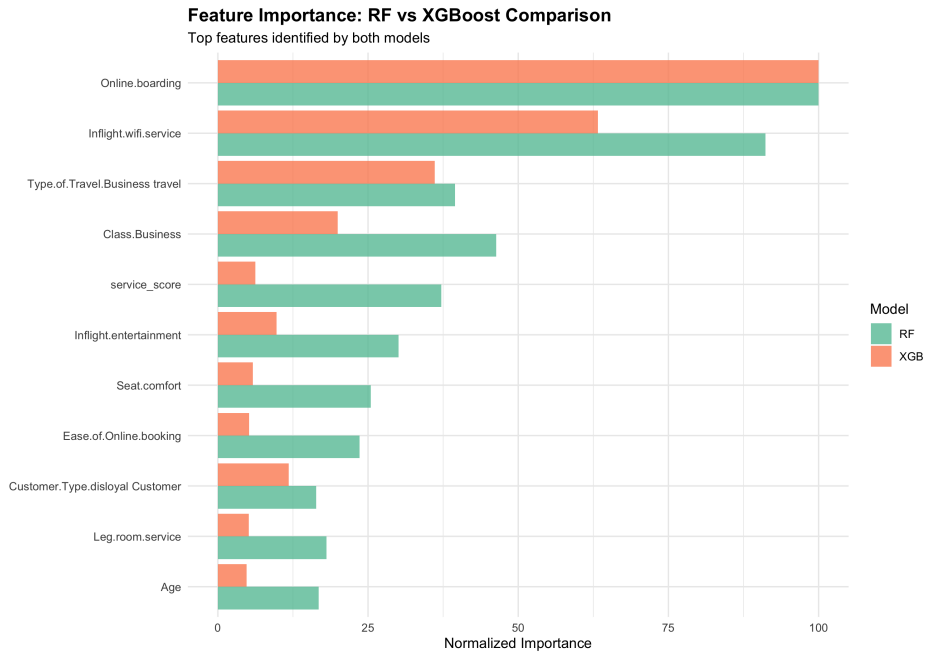


Figure 22: Figure 5.6 Feature importance: RF vs XGBoost comparison

3.5.4 How do key features shape satisfaction? (Partial dependence) To better interpret the effect of the most influential predictors, we generate a **partial dependence plot (PDP)** for the top feature (**Online boarding**) using the XGBoost model (Figure 5.7). Holding other variables constant, predicted satisfaction probability increases **monotonically** with higher online boarding service ratings. The curve shows that as online boarding service quality improves from low to high ratings, the predicted probability of satisfaction increases substantially, indicating that this service dimension has a strong direct effect on passenger satisfaction.

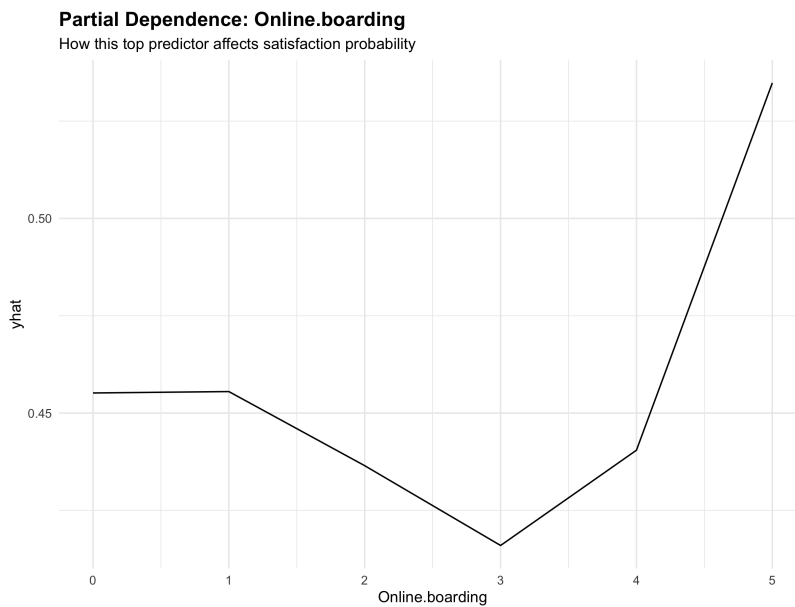


Figure 23: Figure 5.7 Partial dependence plot for Online boarding (top predictor)

The partial dependence plot demonstrates that **Online boarding service** has a strong, positive relationship with satisfaction probability. Passengers who rate online boarding highly are substantially more likely to be satisfied, even when controlling for all other service dimensions and passenger characteristics. This finding highlights the importance of digital service touchpoints in the modern airline experience.

3.5.5 Conclusion for RQ5 In summary, the modelling results provide a clear answer to RQ5:

- **Predictive accuracy:** Using tree-based methods (Random Forest and XGBoost), we can **accurately predict passenger satisfaction**, achieving high accuracy (96.2%) and AUC (0.994) on a held-out test set and consistent performance under cross-validation. This suggests that airlines can use routinely collected survey and operations data to anticipate satisfaction outcomes with good reliability.
- **Key predictors:** The most important drivers of predicted satisfaction are:
 1. **Online boarding service**, which emerges as the single most critical predictor (normalized importance: 100.0).
 2. **Inflight wifi service**, ranking second in importance (77.2).
 3. **Travel context** (business vs. personal travel, business vs. economy class), which significantly influences satisfaction expectations.
 4. **Other service dimensions** such as inflight entertainment, seat comfort, and the composite service score.

These findings reinforce that **satisfaction is shaped primarily by service quality, with digital touchpoints (online boarding, wifi) being particularly critical**. For managers, the models can be used not only to forecast satisfaction, but also to prioritize interventions—for example, investing in online boarding infrastructure and inflight connectivity, or customizing service offers for different customer segments and travel purposes.

4. Insight Gained

Through our comprehensive analysis of airline passenger satisfaction data, we have gained several key insights that have important implications for airline operations and service design:

4.1 The preferences for service attributes among different age groups show a high degree of differentiation.

The research results show that younger passengers pay more attention to digital touchpoints, such as online check-in and in-flight Wi-Fi, while older passengers attach more importance to seat comfort and in-flight manual services. This indicates that when formulating service strategies, airlines need to implement differentiated operations for different age groups in order to more effectively enhance the experience.

4.2 Flight distance has changed the mechanism by which passengers evaluate services.

The data indicates that the dining score decreases as the flight distance increases after controlling for other variables. This means that long-haul passengers will not increase their ratings due to more usage opportunities; instead, higher expectations will lead to stricter evaluations.

4.3 The formation logic of satisfaction between business travelers and individual travelers is completely different.

Business travelers are highly sensitive to time efficiency and in-flight comfort, with comfort having a significant impact on their satisfaction. The overall satisfaction of individual travelers is relatively low, and their response to changes in time and comfort is limited. This significant difference indicates that airlines should distinguish between two types of passengers when investing resources: optimizing legroom and time reliability for business passengers may bring significant benefits, while for individual passengers, a differentiated strategy should be adopted in terms of price, entertainment or overall leisure experience.

4.4 The interaction between cabin class and flight duration creates a strong difference in experience.

Both the heat map and ANOVA results show that the satisfaction of economy class passengers on long flight segments has significantly declined, while business class passengers are almost unaffected by the flight distance. Physical attributes such as seat comfort and legroom demonstrate extremely significant importance in the interaction of “cabin size \times range”. This indicates that service improvement should not be confined to a single cabin class or a single flight, but should focus on combinations that are most prone to experience gaps, such as long-haul economy class.

4.5 Service quality has a greater impact on satisfaction than operational indicators.

Although operational variables such as delays do have an impact on the experience, their importance in the model is significantly lower than that of service ratings. This indicates that when passengers evaluate their satisfaction, they pay more attention to the service quality of airlines rather than occasional operational disturbances.

5. Project Adjustment & Lesson Learned

In the early stage of the project, we gradually narrowed the research questions from broad topics to a structure suitable for the existing data analysis, such as focusing on age differences, differences in passenger types, and the interaction between cabin classes and flight routes. This adjustment enhances the interpretability of the model and also makes the conclusion more targeted. Statistical models help us understand the mechanism relationships among variables, such as significant differences, interaction effects and heterogeneity. Machine learning models, on the other hand, have stronger capabilities in prediction and feature ranking. By reasonably adopting two models in the project, we enhanced the completeness of the research. Team members have also formed a good division of labor and collaboration, making the entire research process smoother and helping us correct the analysis direction in a timely manner.

6. Team Collaboration

The team made clear divisions of labor based on their respective research questions. Clear task division reduces repetitive work, enabling the team to advance multiple stages in parallel. Meanwhile, we ensured the reproducibility of all work by sharing code and visualizing results, thereby enhancing the overall project quality. The team maintains weekly discussions throughout the project cycle to solve the problems they encounter in their research through timely discussions. This communication mechanism enables members to solve problems in a timely manner and update analytical methods when necessary. Overall, teamwork not only enhanced efficiency but also contributed to the consistency of the final report.

7. Future Directions

In the future, the dynamic characteristics of satisfaction can be studied by incorporating route categories or seasonal changes. In addition, if free text comments from passengers can be obtained, the dissatisfaction points and potential needs of passengers can be further understood through sentiment analysis. As the machine learning model in this study has predictive capabilities, it can be used in the future for personalized service recommendations or operational strategy optimization. For instance, it can predict the risk of dissatisfaction for high-value passengers in advance and take intervention measures, helping airlines make the best service investment decisions within budget constraints.

8. Conclusion

This study, through a systematic analysis of multiple factors such as age, passenger type, cabin class and flight distance, reveals that the formation mechanism of air passenger satisfaction has significant heterogeneity and situational dependence. The results show that digital services play a core role in the modern aviation experience, and seat comfort, legroom and entertainment systems are particularly important in long-haul

flights and economy class. Meanwhile, business travelers pay more attention to time and comfort. Further machine learning models indicate that passenger satisfaction is not only predictable but also mainly driven by service quality rather than operational disturbance factors. Overall, this study emphasizes that airlines should focus on differentiated services and digital services in their strategies to enhance satisfaction, thereby achieving more effective experience management and resource allocation in the face of diverse passenger demands.