

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

2025-12-05

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 departure and arrival delays influence passenger satisfaction differently between loyal and disloyal customers?
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:

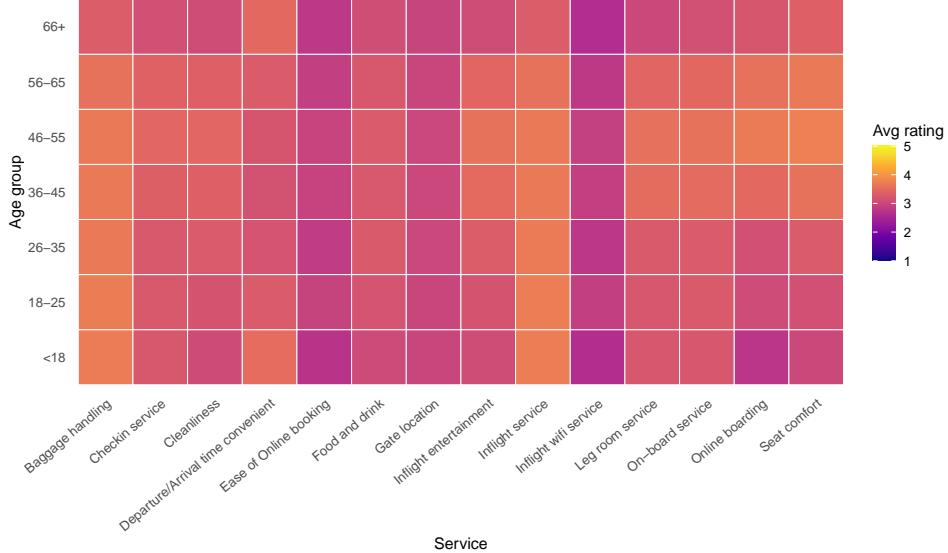


Figure 1: Figure 3.1 Average service ratings by age group

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:

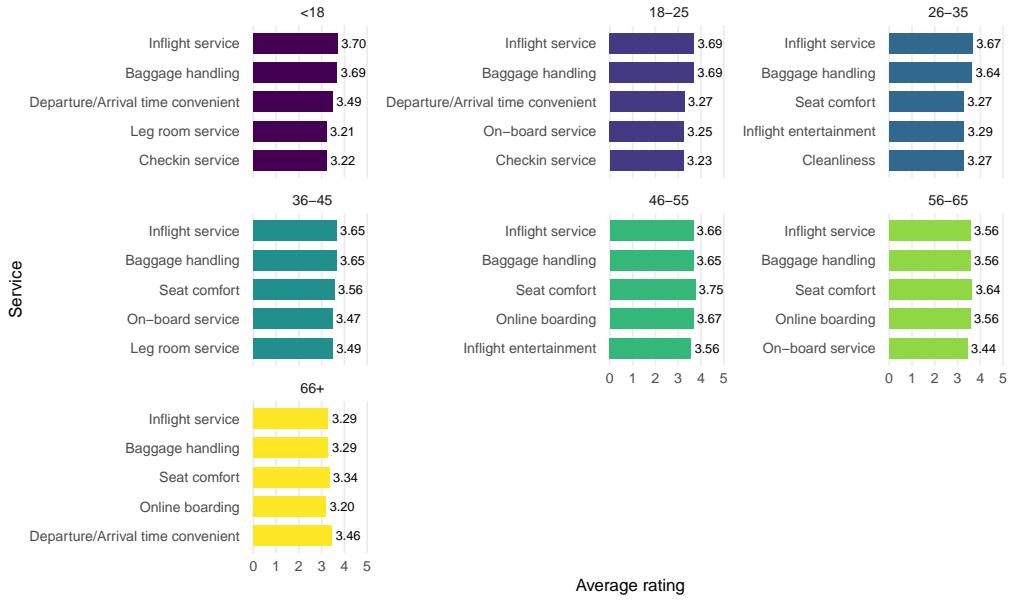


Figure 2: Figure 3.2 Top-5 services by age group

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.

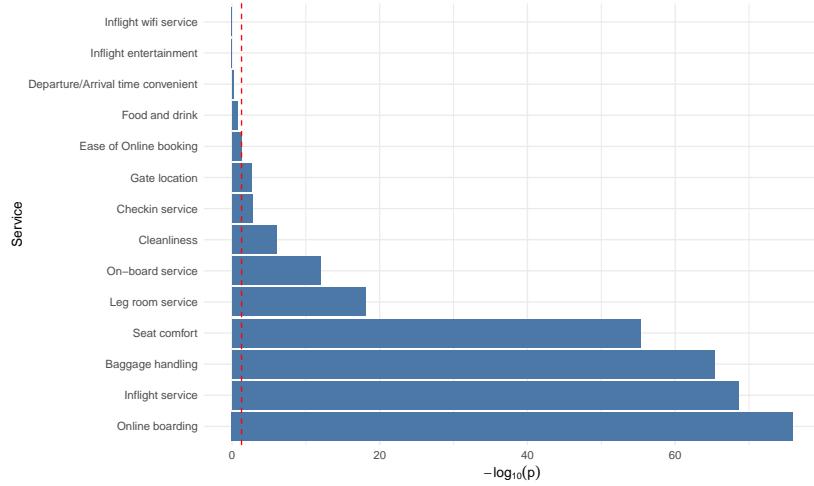


Figure 3: Figure 3.3 K–S test results

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:

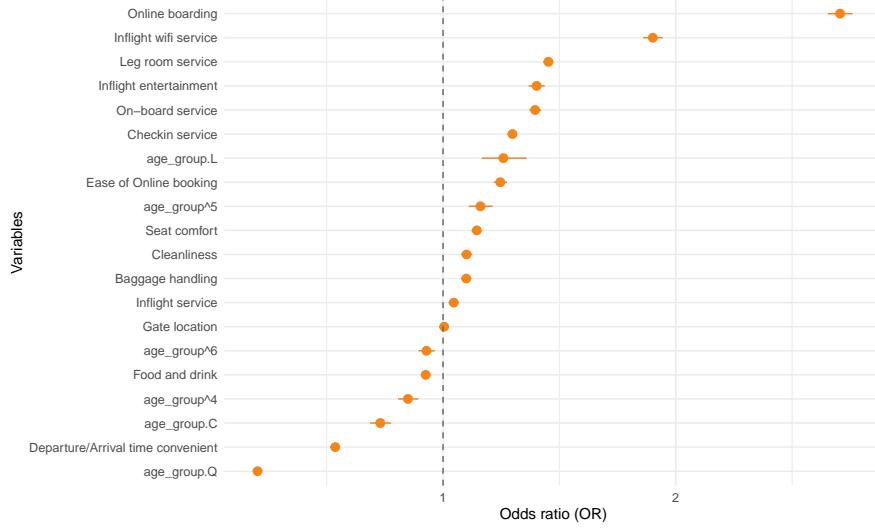


Figure 4: Figure 3.4 Logistic regression

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?

3.3 How do departure and arrival delays influence passenger satisfaction differently between loyal and disloyal customers?

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):

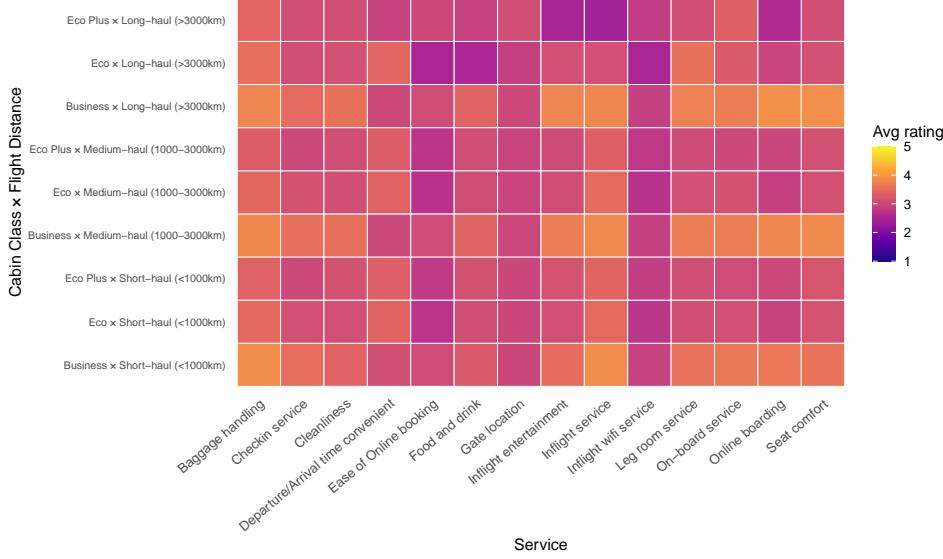


Figure 5: Figure 3.5 Heatmap of Average Ratings

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:

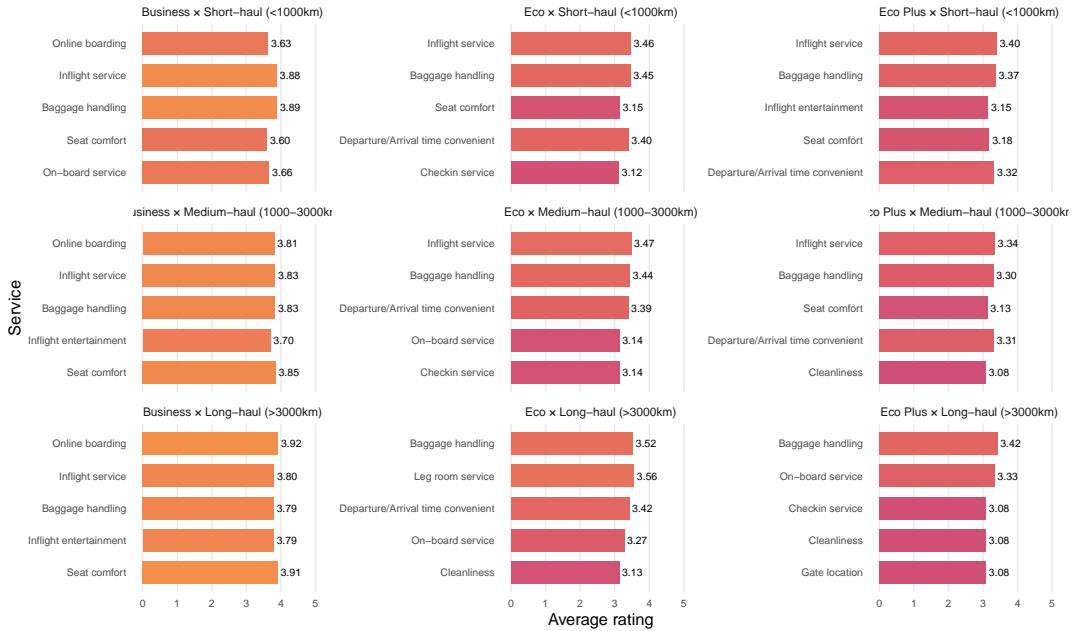


Figure 6: Figure 3.6 Top-5 Services by Cabin Class × Flight Distance

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:

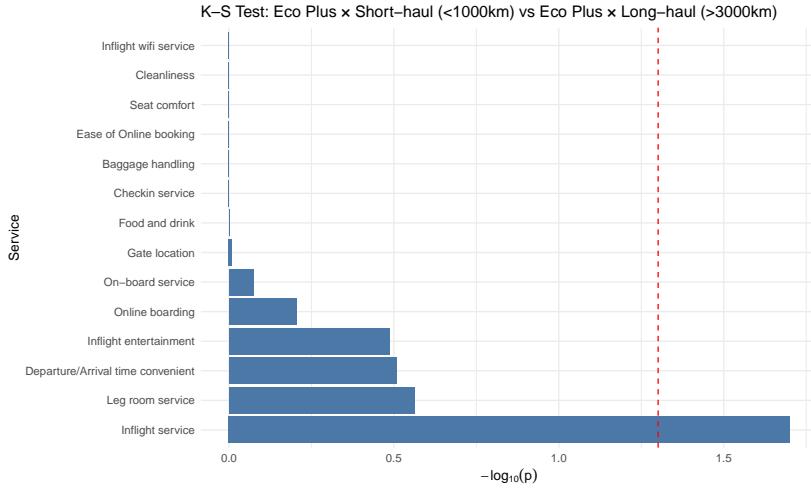


Figure 7: Figure 3.7 K–S Test Results (Eco Plus Short vs. Long)

The results show that “Inflight service” has the most statistically significant distributional difference ($-\log_{10}(p)$ value) 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:

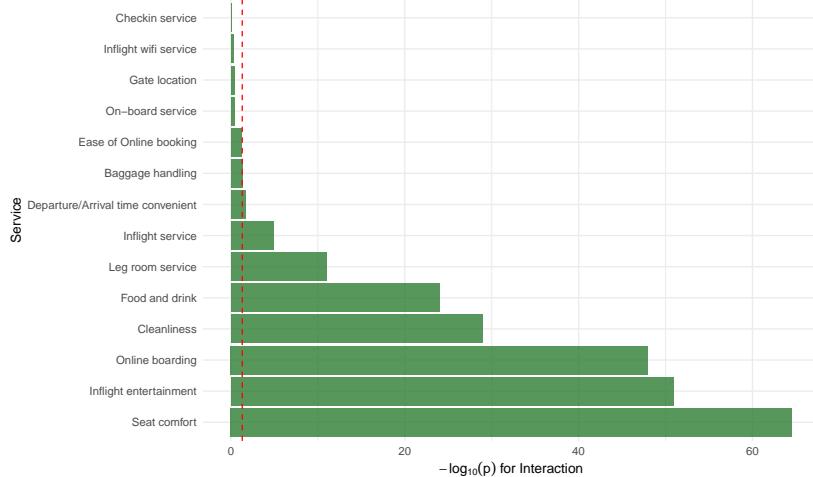


Figure 8: Figure 3.8 ANOVA Interaction Effects

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%).

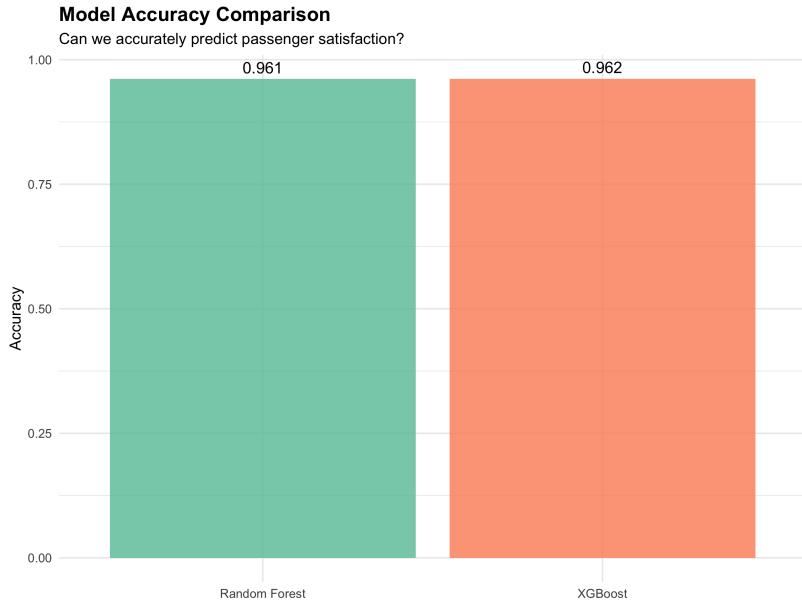


Figure 9: Figure 5.1 Model accuracy comparison

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 10: 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.

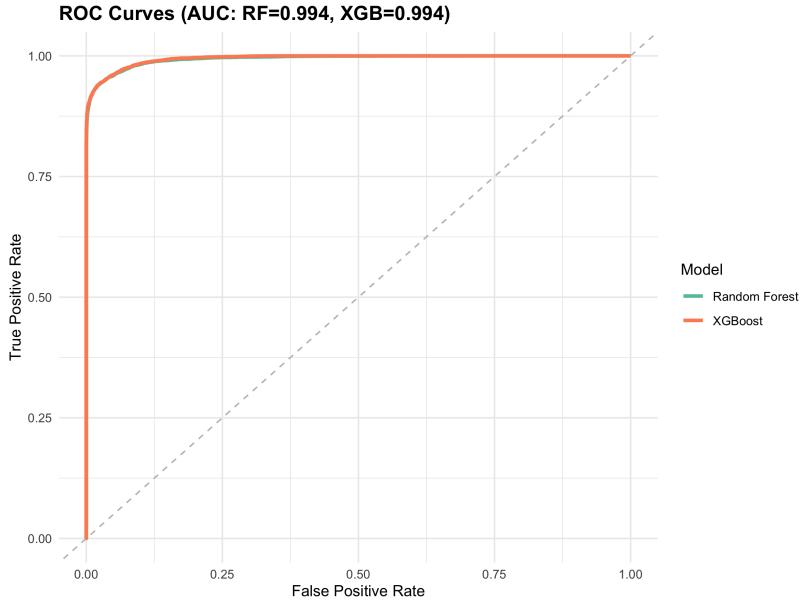


Figure 11: Figure 5.3 ROC curves comparison

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.

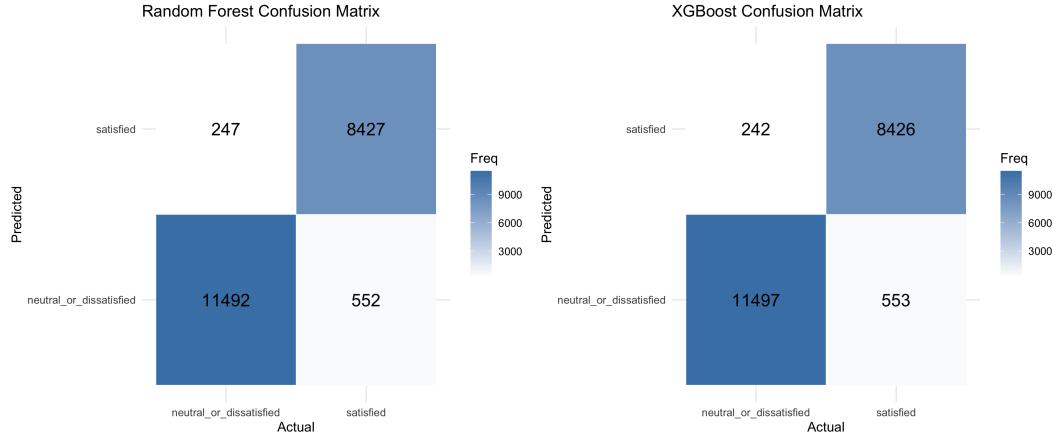


Figure 12: Figure 5.4 Confusion matrices for Random Forest and XGBoost

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.

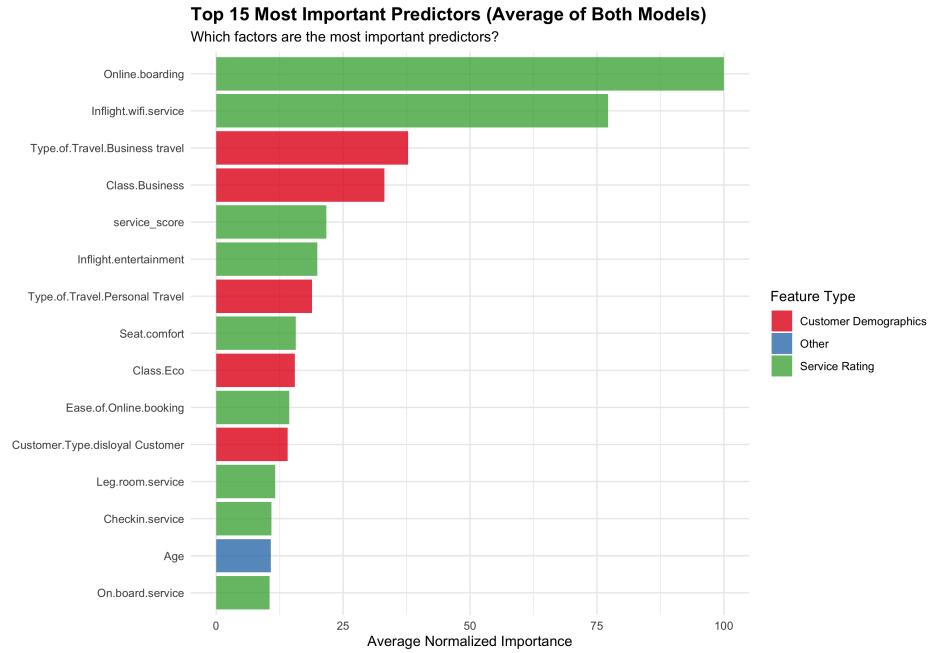


Figure 13: Figure 5.5 Top 15 most important predictors (average of both models)

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.

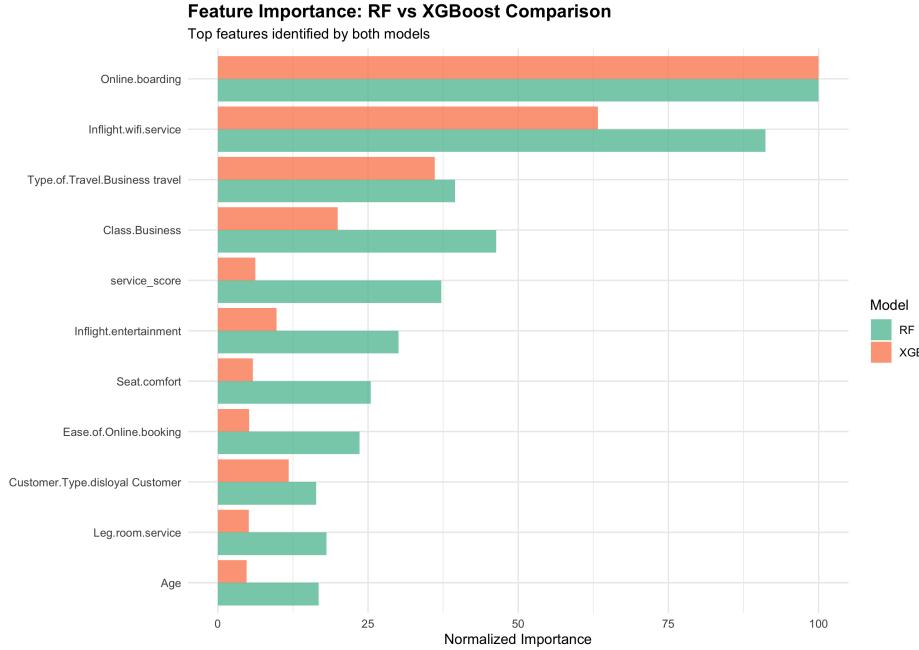


Figure 14: Figure 5.6 Feature importance: RF vs XGBoost comparison

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.

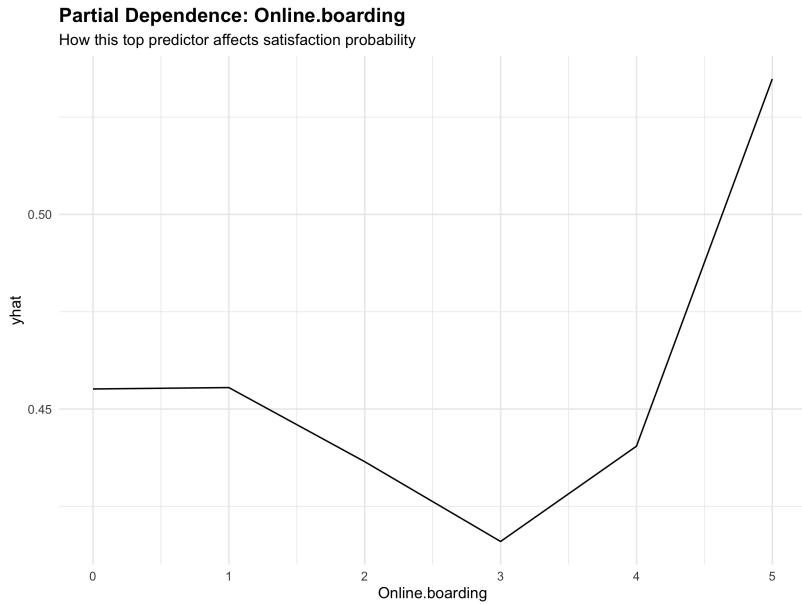


Figure 15: Figure 5.7 Partial dependence plot for Online boarding (top predictor)

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**
5. **Project Adjustment & Lesson Learned**
6. **Team Collaboration**
7. **Future Directions**
8. **Conclusion**