Airline Passenger Satisfaction Analysis

Mohd Saif (C837319113)
Mehar Bawa (C837448655)
Tharunya Pathipati (C837228323)

College of Business, Colorado State University

CIS 575: Applied Data Mining and Analytics in Business

Dr. Gaurav Jetley

December 05, 2025

Table of Contents

1. Executive Summary

Passenger satisfaction is a critical performance indicator for airlines, influencing repeat purchases, customer loyalty, operational reputation, and competitive positioning. As airlines balance increasing passenger volumes with operational constraints and rising expectations, understanding the drivers of satisfaction becomes essential for targeted improvement efforts.

This project applies data mining techniques to a publicly available airline passenger satisfaction dataset that includes demographic information, trip characteristics, operational delay metrics, and detailed service ratings. Our objective is to build predictive models capable of classifying passengers as "satisfied" or "neutral/dissatisfied," while also identifying the most influential factors contributing to satisfaction outcomes.

The project follows a structured life cycle: data selection, data preparation, exploratory analysis, model training, and model comparison. While the dataset contains a broad mixture of numeric and categorical variables, preliminary findings indicate that both operational performance (e.g., delays) and service quality metrics (e.g., seat comfort, online boarding, inflight entertainment) potentially play significant roles.

This deliverable includes four predictive models: Decision Tree, Logistic Regression, Random Forest and Neural Network, evaluated using standard performance measures. Using these results, we will provide actionable recommendations on how airlines can optimize operations, prioritize service improvements, and develop proactive strategies to reduce dissatisfaction. The insights gained from this analysis can guide future resource allocation and enhance overall passenger experience.

2. Business Problem

Airlines operate in an environment characterized by high competition, tight margins, and customers with increasingly high expectations. Dissatisfied passengers may switch carriers, leave negative reviews, or decrease their customer engagement, directly affecting revenue and brand perception.

Despite collecting extensive operational and survey data, many airlines lack a data-driven framework to understand which factors most strongly influence satisfaction. Key questions remain unanswered:

- Which aspects of the passenger journey, such as service quality like inflight Wi-Fi, cleanliness, Food and drink, delays, or online processes; most affect satisfaction?

- How do demographic or trip characteristics influence satisfaction patterns?

- Can we proactively predict which passengers are at risk of dissatisfaction before the flight experience concludes?

Without predictive insights, service improvements may be misaligned, operational decisions may remain reactive, and opportunities to mitigate dissatisfaction may be missed.
This project addresses the core business need to transform airline data into predictive insights that support strategic and operational decision-making.

3. Business Objectives

The goal of the project is to use data mining methods to create a comprehensive understanding of passenger satisfaction and develop predictive capabilities that enable proactive actions. The specific objectives are:

3.1. Identify Key Drivers of Passenger Satisfaction

Determine which operational, demographic, and service-related factors contribute most significantly to passenger satisfaction. This includes analyzing delay patterns, travel class, flight distance, and detailed in-flight service ratings.

3.2. Develop Predictive Models

Build classification models (Decision Tree, Logistic Regression, Neural Network, Random Forest) to estimate whether a passenger will be satisfied or neutral/dissatisfied. These models will enable the airline to anticipate negative experiences before they occur.

3.3. Compare Operational vs. Service Quality Influences

Assess whether operational metrics (e.g., departure delay, arrival delay) or service quality scores (e.g., seat comfort, online boarding, inflight entertainment) have stronger effects on satisfaction outcomes.

3.4. Generate Actionable Recommendations

Provide insights to help the airline make data-driven decisions, such as:

- Which service dimensions require investment

- Which passenger groups or flight types warrant additional attention

- Which operational bottlenecks most strongly reduce satisfaction

These objectives ensure that the analysis moves beyond descriptive statistics to generate meaningful, actionable business insights.

4. Process Followed for Selecting and Gathering Data

A structured approach was used to select and gather the dataset for this project. The data must be sufficiently large, relevant, and rich in both operational and service-related variables to support predictive modeling.

4.1 Data Source

The dataset used in this project is a well-known Airline Passenger Satisfaction dataset published on Kaggle (https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction). This dataset was selected because:

- It directly aligns with the goal of predicting passenger satisfaction.

- It includes a broad mix of passenger demographic details, flight characteristics, delay information, and service ratings.

- It contains a sufficiently large sample size (129880).

- The variables are appropriate for both exploratory analysis and model development.

4.2 Data Contents

The dataset includes:

- **Target Variable**

    o Satisfaction label (Satisfied vs. Neutral/Dissatisfied)

- **Demographic Features**

    o Gender

    o Age

- **Trip Characteristics**

    o Type of Travel (Business / Personal)

    o Travel Class

    o Flight Distance

- **Operational Metrics**

    o Departure Delay in Minutes

    o Arrival Delay in Minutes

- **Service Quality Ratings**

- Seat Comfort

- Inflight Entertainment

- Online Boarding

- Food and Drink

- On-board Service/Cabin Service

- Baggage Handling

- Cleanliness

- Check-in Service

- Gate Location

5. Preliminary Data Exploration and Findings

Initial exploratory steps were conducted to understand data distribution, identify potential quality issues, and form early hypotheses about satisfaction patterns. These findings provide foundational context for modeling.

5.1 Overview of Target Variable

A preliminary review shows that the dataset contains two main classes:

- **Satisfied**

- **Neutral / Dissatisfied**

The distribution appears slightly imbalanced(Satisfied: 44%, Neutral / Dissatisfied: 56%), with more satisfied passengers than dissatisfied ones. This must be considered during modeling to avoid biased accuracy metrics.

5.2 Observations from Early Exploration

Although detailed statistics will be added later, high-level trends observed during preliminary review include:

- **Demographics**

  - A wide age distribution, with many passengers in mid-age (40-60) groups.

  - No immediate extreme skew in gender distribution.

Age Distribution by Satisfaction Level
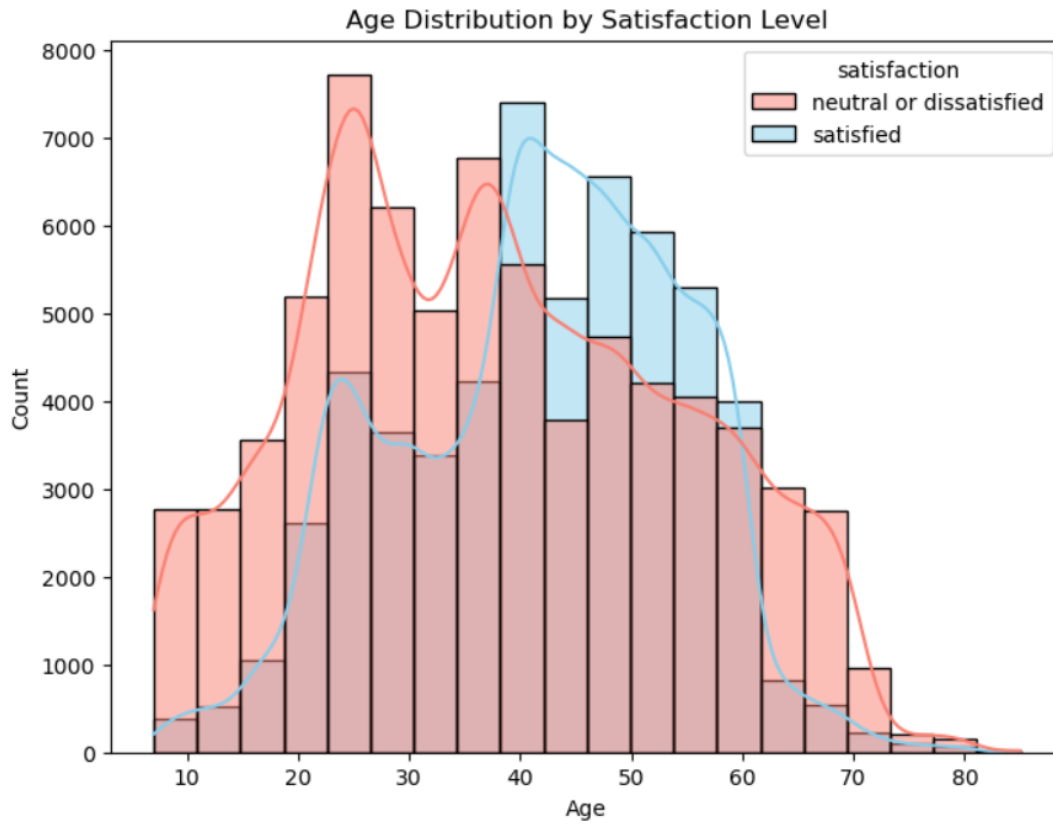
- **Trip Characteristics**
  - Business travelers appear more represented in higher satisfaction categories.
  - Travel class seems to significantly influence satisfaction, with business class trending higher.

Class: Business — Class: Eco
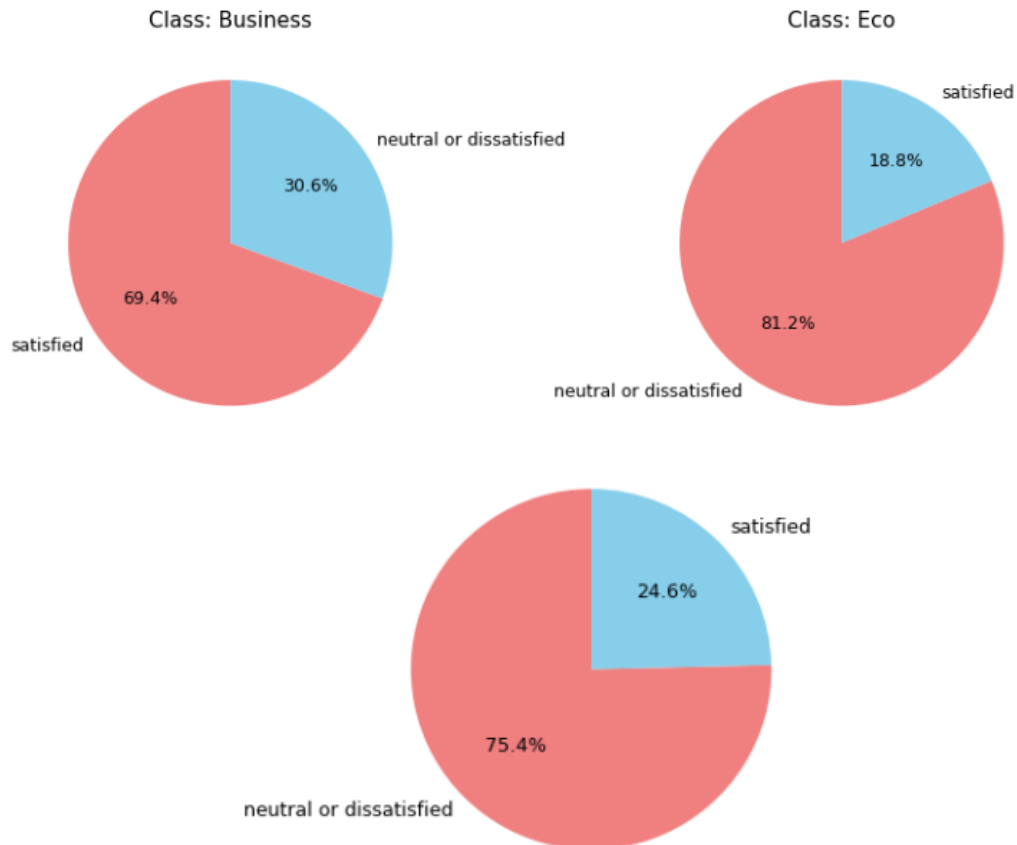
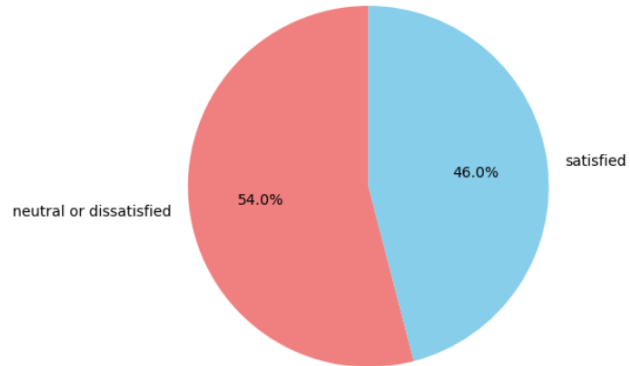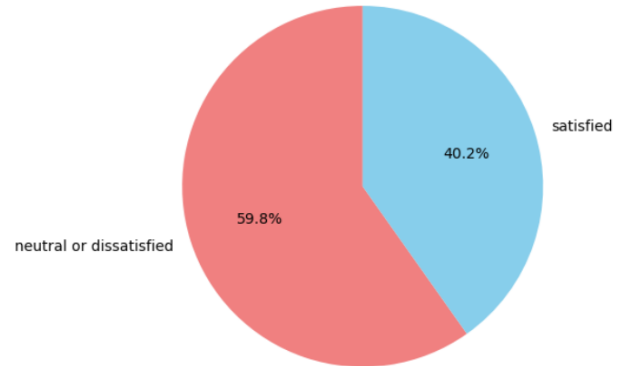- **Operational Variables**

  - A considerable portion of flights have minimal delays.

  - Passengers on delayed flights appear more likely to fall in the neutral/dissatisfied category.

Satisfaction: On-Time Flights (Departure Delay = 0)

satisfied 46.0%

neutral or dissatisfied 54.0%

Satisfaction: Delayed Flights (Departure Delay > 0)

satisfied 40.2%

neutral or dissatisfied 59.8%

Satisfaction: On-Time Flights (Arrival Delay = 0)

satisfied 47.4%

neutral or dissatisfied 52.6%
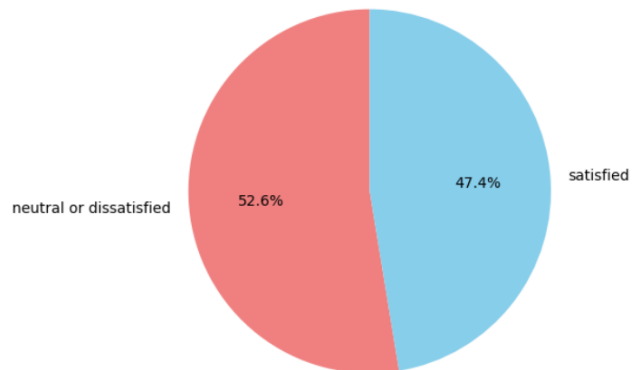
Satisfaction: Delayed Flights (Arrival Delay > 0)

satisfied 38.4%

neutral or dissatisfied 61.6%
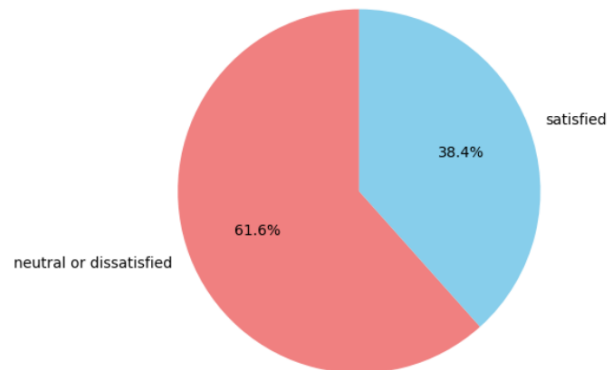
- **Service Ratings**

  o Ratings such as inflight Wi-Fi and inflight entertainment appear to be positive factors and Gate Location as negative.

  o Ratings in these areas tend to affect the satisfaction.
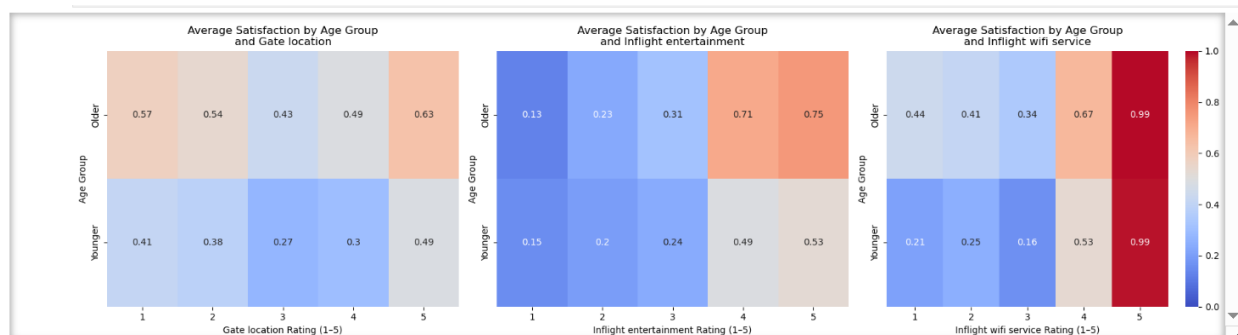
6. Data Preparation

All data preparation was performed in **KNIME Analytics Platform** using a structured workflow that transformed the raw airline passenger satisfaction dataset into modeling-ready tables. The goal of this stage was to ensure data quality, handle mixed data types, and create consistent inputs for all downstream models.

6.1 Data Import and Initial Filtering

The raw dataset was imported using a **CSV Reader** node. A **Row Filter** node was then used to remove clearly invalid or unusable records (e.g., rows with missing target labels or obviously corrupted values like 0 rating in Seat comfort, Gate Location). This ensured that only complete, analyzable passenger records were retained.

6.2 Handling Missing Values

A **Rule Engine - Missing Value** step was applied to identify and treat missing or inconsistent entries. Depending on the field:

- Some values were imputed and Mode values of the column (for categorical ratings like inflight wifi service).

6.3 Feature Selection and Column Filtering

Several **Column Filter** nodes were used to drop fields that were not relevant for prediction or could leak information, such as:

- Pure identifiers (e.g., passenger ID and Sequence No columns).

6.4 Encoding Categorical Variables

Because many predictors were nominal (e.g., gender, travel class, type of travel), **One-to-Many (one-hot) encoding** nodes were used. Each categorical variable was converted into a set of binary indicator columns. This encoding is required for Logistic Regression and Neural Networks and also standardizes inputs for tree-based models.

6.5 Normalization of Numerical Features

A **Normalizer** node was used to scale numeric variables (e.g., age, distance, delay minutes, rating scores). This preprocessing was especially important for **Logistic Regression** and **Neural Networks**, which are sensitive to feature magnitude.

Later, when comparing Logistic Regression with and without scaling, we observed that normalization increased validation accuracy from **72.78%** to **88.06%**, clearly demonstrating the importance of this step for linear models.

6.6 Train-Validation Partitioning

To support fair model assessment, the prepared data was split into **training** and **validation** sets using **Partitioning** nodes (e.g., 70% training / 30% validation). Separate partitioning nodes were used for different modeling branches to align with the specific feature subsets and normalization pipelines. All performance metrics reported in this report (accuracy, recall, etc.) are based on the validation sets.

6.7 Correlation and Exploratory Analysis

Prior to modeling, correlation analysis was performed in Python using Jupyter Notebook to understand relationships between features and satisfaction. The satisfaction associations were typically found between:

- Strongest: Service quality ratings (online boarding, seat comfort, inflight entertainment, inflight wifi)

- Weakest Variables (age, gender, gate location)

These patterns informed expectations going into the modeling phase and were later confirmed by model feature importance outputs.

7. Data Modeling and Assessment

Four main model families were implemented:

1. **Logistic Regression (with and without Normalization)**

2. **Decision Tree (standard and pruned)**

3. **Random Forest**

4. **Neural Network (with and without parameter optimization)**

For each model, KNIME **Scorer** nodes were used to compute confusion matrices and performance metrics such as **accuracy, sensitivity/recall, specificity, and precision**. Because the business goal is to identify **at-risk (neutral/dissatisfied) passengers**, particular attention was paid to **recall for the dissatisfied class**, not only overall accuracy.

7.1 Logistic Regression

Two versions of Logistic Regression were built:

7.1.1 Logistic Regression without Normalization

- **Validation accuracy: 72.78%**

- **Recall: 93.00%**

The initial model was trained directly on unscaled numerical features. Performance was modest, indicating that the differing scales of variables (e.g., delays vs. distance) Gini negatively impacted the stability and effectiveness of the model.

7.1.2 Logistic Regression with Normalization

- **Validation accuracy: 88.06%**

- **Recall: 85.70%**

After applying normalization to all numeric features, the Logistic Regression model improved significantly. Coefficients became more stable, and overall accuracy increased by more than **15 percentage points**.

This experiment clearly demonstrates that **feature scaling is critical** for linear models in this context. The scaled Logistic Regression served as a strong, interpretable baseline, although its performance still lagged behind tree-based ensembles.

7.2 Decision Tree

Two Decision Tree models were built:

7.2.1 Standard Decision Tree

- **Validation accuracy: 94.33%**

- **Recall: 92.60%**

The standard tree produced a clear segmentation of passengers into satisfied vs. dissatisfied groups. Splits of Decision trees involved combinations of:

- Service quality ratings (e.g., low online boarding or seat comfort → dissatisfaction), and

- Delay thresholds (long delays → higher dissatisfaction).

7.2.2 Pruned Decision Tree

- **Validation accuracy: 95.53%**

- **Recall: 93.20%**

A pruned version of the Decision Tree, implemented using a Gini quality index through MDL pruning method and reduced error pruning transformation, removed over-specific branches (from 3235 to 281) and resulted in slightly higher validation accuracy. This model captured the main decision logic with fewer, more interpretable rules.

While the pruned tree had excellent interpretability and strong performance, it was still marginally outperformed by the Random Forest ensemble.

7.3 Random Forest

- **Validation accuracy: 95.76%**

- **Recall: 93.20%**

The **Random Forest** model achieved the best performance of all models tested. By aggregating many decision trees trained on bootstrapped subsets of the data and features, Random Forest:

- Captured complex nonlinear relationships;

- Was robust to noise and overfitting.

Feature importance score (see Appendix 3) from the Random Forest indicated that the most influential predictors included:

- Online boarding

- Seat comfort

- Inflight entertainment

- Inflight wifi

These rankings show that convenience/service quality measures are key determinants of satisfaction.

7.4 Neural Network (Multi-Layer Perceptron)

Two neural network models were developed using an MLP architecture:

7.4.1 Baseline Neural Network

- **Validation accuracy: 90.14%**

- **Recall: 87.90%**

The baseline MLP, using default or manually chosen hyperparameters, produced reasonable performance but did not match the Decision Tree or Random Forest.

7.4.2 Optimized Neural Network

- **Validation accuracy: 91.70%**

- **Recall: 88.20%**

Using KNIME's **Parameter Optimization Loop**, key hyperparameters (e.g., number of hidden units, maximum iterations) were tuned. This improved validation accuracy by about **1.5 percentage points**, yet the Neural Network still fell short of the pruned Decision Tree and Random Forest models.

Neural Networks showed promise but required more effort to tune and were less understandable than tree-based methods.

8. Model Comparison and Selection

Table 1 summarizes the validation accuracies and recall for all models:

Model	Validation Accuracy	Recall
Logistic Regression (unscaled)	72.78%	93.00%
Logistic Regression (normalized)	88.06%	85.70%
Decision Tree	94.33%	92.60%
Pruned Decision Tree	95.53%	93.20%
Random Forest	95.76%	93.20%
Neural Network (no optimization)	90.14%	87.90%
Neural Network (with optimization)	91.70%	88.20%

8.1 Accuracy and Class-Imbalance Considerations

The dataset is slightly imbalanced, with more "Satisfied" than "Neutral/Dissatisfied" passengers. For this reason, **accuracy alone is not sufficient** to evaluate models.

In addition to accuracy, we examined:

- **Sensitivity/Recall for the Neutral/Dissatisfied class** (how many unhappy passengers we correctly flag),

- **Precision** for the Neutral/Dissatisfied class (of those we flag as unhappy, how many truly are), and

- **F1 score**, which balances precision and recall.

From a business perspective, **missing a dissatisfied passenger (false negative)** is more costly than mistakenly flagging a satisfied one (false positive). Thus, **recall for the dissatisfied class** was treated as the most important performance measure, and accuracy was used as a supporting metric.

Across the models, those with higher overall accuracy (Random Forest and the pruned Decision Tree) also showed strong recall for the dissatisfied class in their confusion matrices, while the unscaled Logistic Regression performed poorly both in accuracy and recall.

8.2 Final Model Selection

Given all evaluation criteria, **Random Forest** was selected as the **final model** for this project because:

- It achieved the **highest validation accuracy (95.76%).**

- It maintained strong performance for the dissatisfied class (good recall and F1 compared to other models).

- It has the largest ROC-AUC Curve Area of 0.993 (See Appendix 4).

- It is robust to noise and capable of modeling complex interactions.

- It provides intuitive feature importance measures that support business interpretation using graphs.

The pruned Decision Tree served as a close second choice. However, Random Forest offers a slightly stronger and more stable predictive performance and is therefore recommended as the primary operational model.

9. Conclusions and Recommendations

9.1 Key Analytical Conclusions

1. **Random Forest is the best-performing model**

   o With a validation accuracy of **95.76%**, Random Forest outperformed Logistic Regression, Decision Tree, and Neural Network models.

   o It also offered a strong recall of **93.20%** for the **Neutral/Dissatisfied** class, aligning with the business objective of identifying at-risk passengers.

2. **Data preprocessing, especially normalization, is critical**

   o Logistic Regression accuracy increased from **72.78%** to **88.06%** after normalization, highlighting the importance of proper scaling in linear models.

3. **Service quality variables are major drivers of satisfaction**

   o Feature importance rankings consistently highlighted **online boarding, seat comfort, inflight entertainment, and inflight wifi** as key predictors. These service factors were more influential than demographic variables.

4. **Operational delays still matter**

o Arrival and departure delays were also important predictors, particularly for identifying dissatisfied passengers, reinforcing punctuality in customer experience.

9.2 Business Recommendations

Based on the modeling results and feature importance analysis, the airline can consider the following actions:

9.2.1 Enhance High-Impact Service Dimensions

- **Online boarding:** Simplify and improve mobile/web check-in workflows, reduce technical issues, and provide clearer instructions.

- **Seat comfort:** Evaluate opportunities for improved seating, especially on longer routes where discomfort is more pronounced.

- **Inflight wifi & entertainment:** Ensure robust, up-to-date content and reliable connectivity to the internet is available.

Investments in these areas are likely to yield measurable increases in overall satisfaction.

9.2.2 Monitor and Mitigate Delay-Related Dissatisfaction

Even minor operational improvements can have noticeable impact:

- Optimize turnaround processes and gate operations to reduce delays.

- When delays are unavoidable, emphasize **timely communication** and **proactive support** which may soften their impact on satisfaction.

9.2.3 Deploy a Predictive "At-Risk Passenger" Alert System

Using the Random Forest model, the airline can:

- Predict which passengers on a given flight are most likely to be dissatisfied based on their profile, service ratings, and operational conditions;

- Flag high-risk flights or segments;

- Trigger targeted interventions (e.g., personalized messages, vouchers, seat changes, additional attention from cabin crew).

This shifts the airline from reactive complaint handling to **proactive experience management**.

9.3 Limitations and Future Work

- The analysis relies on a publicly available dataset; real airline data may contain additional fields (e.g., loyalty tier, route, aircraft type) that could further improve the model.

- Only a subset of algorithms was tested. Future work could explore **gradient boosting methods** (e.g., XGBoost, AdaBoost) and advanced explainability tools such as **SHAP** for more detailed insight into individual predictions.

- A route- or segment-specific analysis (e.g., long-haul vs. short-haul, business vs. leisure routes) could provide more granular recommendations.

*9.4 What did you learn from the analysis*

The analysis provided clear insight into the factors that most strongly influence airline passenger satisfaction. We learned that service-quality variables, such as online boarding, seat comfort, inflight entertainment, and inflight Wi-Fi, play a significantly larger role in predicting satisfaction than demographic variables. We also observed that operational delays, while secondary to service quality, still meaningfully affect dissatisfaction. The correlation heatmaps and model feature importance outputs (Appendix 1 & 2) reinforced these findings, confirming that improvements in both convenience and service features would have the greatest positive impact on customer satisfaction.

*9.5 How the Results Address the Business Need*

Our predictive modeling objectives were fully met. We successfully built, tested, and compared multiple models, including Logistic Regression, Decision Trees, Random Forest, and Neural Networks, and evaluated them using both accuracy and recall for the satisfied class.

The results directly address the airline's business problem by enabling a data-driven approach to understanding satisfaction drivers and predicting dissatisfaction before it occurs. The insights reveal where operational and service improvements should be focused, while the predictive model can be deployed as an early-warning system to identify passengers most likely to be dissatisfied. Ultimately, the analysis supports informed decision-making that can enhance loyalty, reduce negative feedback, and strengthen the airline's competitive position.
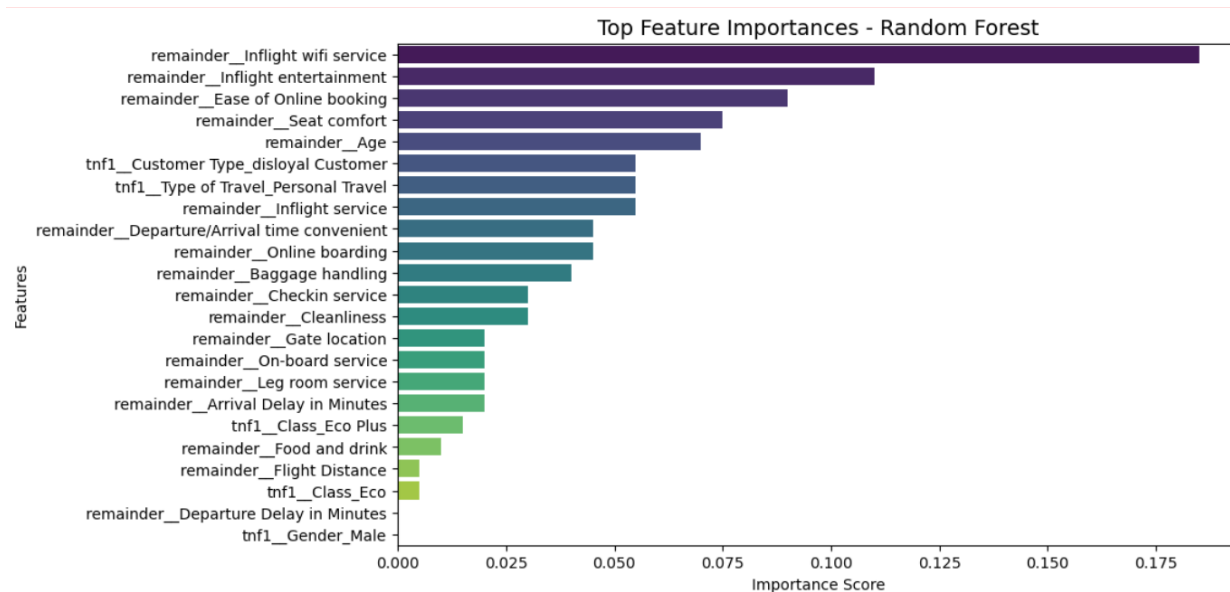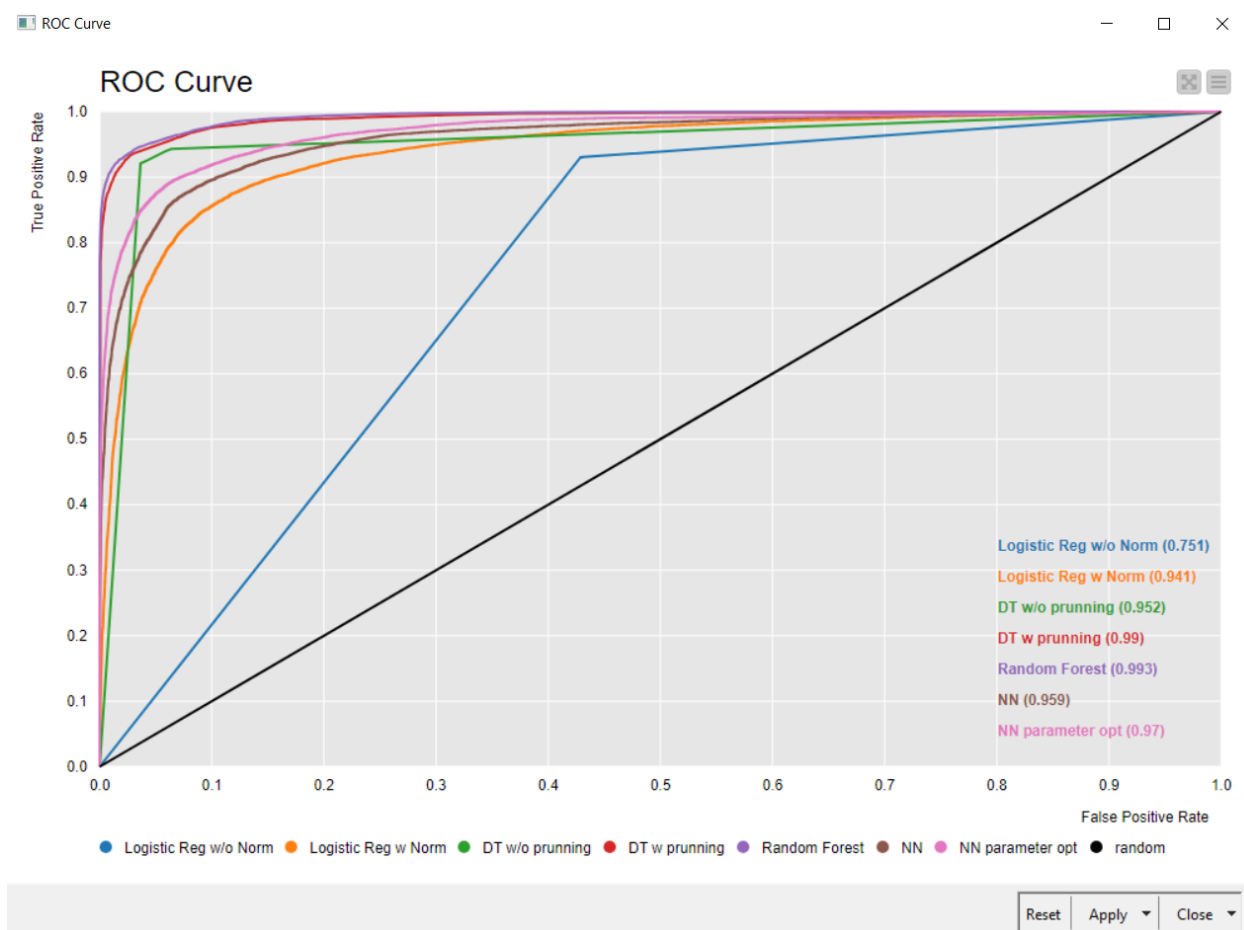
Appendices
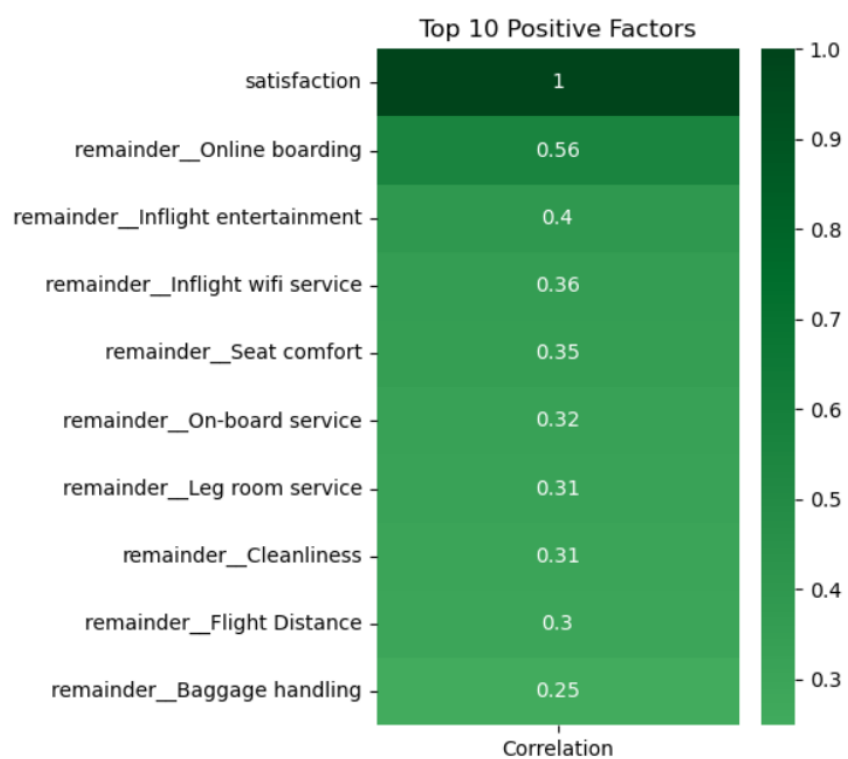
Appendix 1: **Factors positively contributing to Satisfaction**
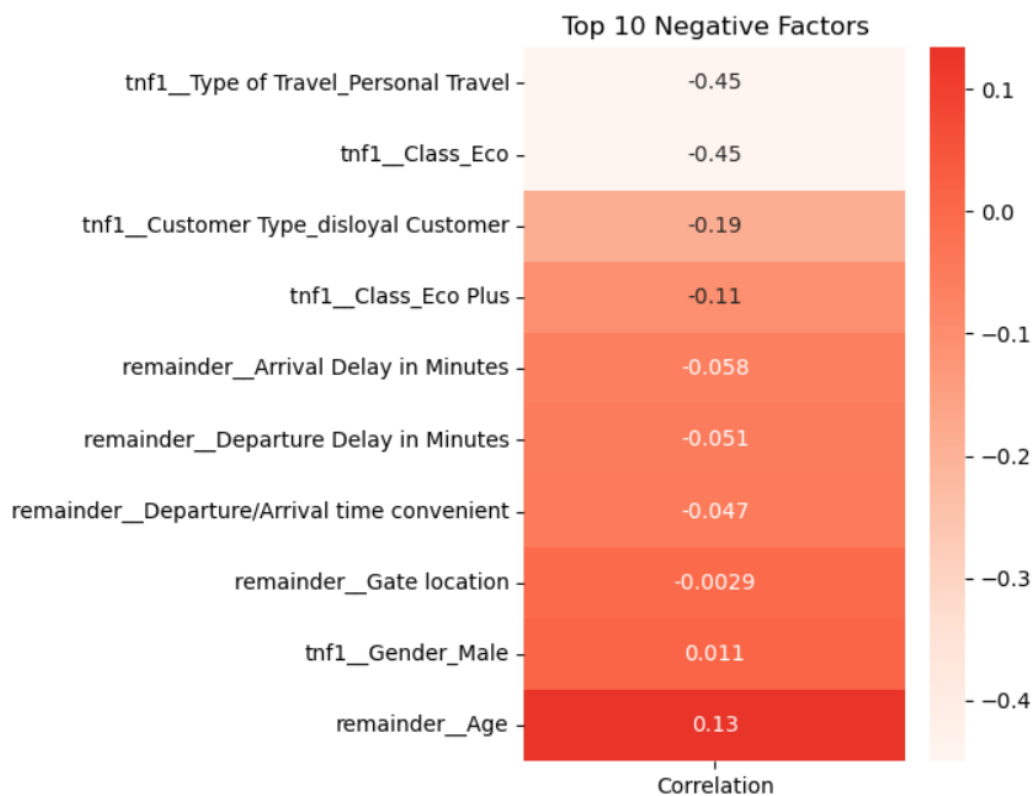
Appendix 2: Factors negatively contributing to Satisfaction

Appendix 3: Fetaure Importance

Appendix 4: ROC-AUC of all Models