Flight to Satisfaction:

Unravelling Customer Satisfaction In The Airline Industry

Search Destination   Find Flight

Project Group 11: Halima Salis, Alero Toki, Rumbidzai Nyoni, Vivian Gunda, and Danisha Goodman

Section I: Background

✈ Search Destination   Satisfaction

In the ever-evolving landscape of the aviation sector, the importance of prioritizing customer satisfaction stands as the basis for success. With customers forming the bedrock of this industry, airlines rely greatly on their patronage. It is, therefore, crucial for organizations to delve deeply into the preferences and desires of their clientele, striving to deliver not just satisfactory but exceptional and tailored experiences. Deploying initiatives such as post-service emails and personalized interactions is one avenue to achieve this goal, fostering a sense of value among customers.

A notable illustration of this customer-centric approach is exemplified by Emirates through their "Knowledge-driven Inflight Service." By empowering their crew to analyze the travel history of customers with the airline, Emirates seeks to enhance the potential for a more personalized and enriching experience for each passenger.

The aviation industry is characterized by intense competition and demands a strategic focus on customer satisfaction, which serves as an important metric for evaluating business performance. The International Air Transport Association (IATA) reports a consistent annual increase of approximately 7% in the number of airline passengers since 2015. Despite this growth, the net profit per airline passenger has witnessed a concerning decline – plummeting from $10 in 2015 to an estimated $7.4 in 2018, with successive decreases to $9 in 2016 and 2017. This decline is primarily attributed to the intense competition within the industry and the recent surge in operational costs faced by airlines. Navigating this challenging landscape requires a renewed commitment to understanding and meeting customer expectations to ensure sustainable success in the aviation sector.

Section II: The Business Problem

Search Destination   Satisfaction

FLIGHT TO
SATISFACTION

Business Problem:
How can airlines improve customer satisfaction?

Problem Statement:
In the fiercely competitive airline industry facing declining profits and escalating operational costs, the critical challenge is to identify and enhance key aspects of customer satisfaction needed to attract and retain passengers.

Analyze:
By exploring the dataset encompassing attributes like cleanliness, seat comfort, service, and overall flight experience, we aim to analyze customer feedback to discern patterns and identify factors significantly impacting satisfaction levels.

Predict:
Utilizing advanced predictive modeling, our objective is to forecast future customer satisfaction, enabling airlines to proactively address potential pain points and strategically allocate resources to areas with the most substantial impact on overall passenger contentment.

Assessment:
Through a comprehensive assessment of the dataset, we seek to not only uncover the primary drivers of customer satisfaction but also provide actionable insights for airlines to emphasize and improve specific aspects of their services, ultimately fostering a more satisfying and personalized flight experience for passengers.

Dataset: https://www.kaggle.com/code/frixinglife/airline-passenger-satisfaction-part-1

Section III: Project Objectives

Search Destination   Satisfaction

Project Objectives

Comprehensive Analysis of Customer Feedback:

Conduct an in-depth examination of the dataset, focusing on parameters such as cleanliness, comfort, service, and overall flight experience.

Identification of Satisfaction Drivers:

Dissect customer feedback to discern patterns and identify the key factors significantly impacting satisfaction levels.

Advanced Predictive Modeling:

Utilize advanced predictive modeling techniques to forecast future customer satisfaction, enabling proactive issue resolution and resource allocation.

Strategic Resource Allocation:

Provide actionable insights for airlines to strategically allocate resources, emphasizing areas with the most substantial impact on overall passenger contentment.

Enhancement of Specific Service Aspects:

Offer a comprehensive assessment of the dataset to uncover primary drivers of customer satisfaction, guiding airlines to emphasize and improve specific aspects of their services.

FLIGHT TO SATISFACTION

Section IV: Primary Goal & Business Impact

✈ Search Destination   Satisfaction

FLIGHT TO
SATISFACTION

Primary Goal: Enhancing Customer Satisfaction

Our project's primary aim is to assist airline businesses in enhancing customer satisfaction. We aim to achieve this by deciphering the elements that resonate with customers, including factors such as cleanliness, comfort, service, and the overall flight experience.

Business Impact: Strategic Insights for Improved Satisfaction

Through the application of predictive modeling on the dataset, our goal extends beyond merely anticipating future customer satisfaction. We aspire to provide actionable and strategic insights to guide airlines on precisely which aspects of their services to emphasize. By fostering an environment that strategically addresses these key factors, our project aims to contribute to increased satisfaction among passengers, ultimately benefiting the business as a whole.

Section V: Project Implications

🛫 Search Destination   Satisfaction

Project Implications

🚀 **Enhanced Customer Loyalty and Retention:** Successful insights implementation leads to increased loyalty and retention.

🚀 **Positive Brand Image and Reputation:** Improved satisfaction contributes to a positive brand image, enhancing the airline's reputation.

🚀 **Competitive Advantage:** Proactive satisfaction improvement provides a competitive edge in the market.

🚀 **Revenue Growth:** Satisfied customers result in word-of-mouth recommendations, driving revenue.

🚀 **Operational Efficiency:** Focused improvements may optimize resources, reducing costs while maintaining service quality.

🚀 **Regulatory Compliance:** Aligning with customer expectations enhances compliance and fosters positive regulatory relationships.

🚀 **Long-Term Sustainability:** Adapting to customer preferences ensures long-term sustainability and industry relevance.

FLIGHT TO SATISFACTION

Section VI: Dataset Description

Search Destination   Satisfaction

FLIGHT TO SATISFACTION

Dataset Overview:

🚀 The dataset was obtained from kaggle and comprises 25 columns and 25976 rows, presenting a comprehensive survey on air passenger satisfaction.

Key Statistics:

🚀 Rows: 25,976

🚀 Columns: 25

Variable/Data Types:

🚀 **Numerical and Categorical:**

✓ Numerical variables include metrics like age, flight distance, delay times etc.

✓ Categorical variables encompass gender, customer type, type of travel, class, and satisfaction level.

Target Variable:

🚀 **Satisfaction:**

✓ Binary classification: Satisfied and Neutral or Dissatisfied.

Classification Problem:

🚀 The primary objective is to predict the passenger's satisfaction level. The model will categorize passengers into one of two levels: Satisfied or Neutral/ Dissatisfied.

Objective:

🚀 Develop a predictive model to anticipate and classify passenger satisfaction levels. This information will empower airlines to address concerns and enhance overall customer satisfaction proactively.

Section VII:
Data Gathering

✈ Search Destination   Satisfaction

Data Gathering:
Understanding Passenger Satisfaction

Dataset Origin

✓ The dataset comprises diverse information collected from an anonymous airline's website.

✓ It includes parameters such as cleanliness, comfort, service quality, and overall flight experience.

Data Source

✓ The data is sourced from a collection mechanism implemented by the airline.

✓ Insights are drawn from post-flight surveys, feedback forms, and customer reviews.

Data Scope

✓ The dataset spans a significant number of observations (25976 rows).

✓ Each row represents a unique passenger experience with associated variables.

Data Gathering:
Understanding Passenger Satisfaction

FLIGHT TO SATISFACTION

Data Types

- ✓ The dataset incorporates a mix of numerical and categorical variables.
- ✓ Numerical variables include flight distance, delay times, and service ratings.
- ✓ Categorical variables encompass gender, customer type, travel type, and satisfaction levels.

Data Quality

- ✓ Prior to analysis, missing values were addressed through imputation.
- ✓ The dataset is now complete, ensuring the reliability of insights.

Temporal Aspect

- ✓ The dataset may include temporal elements, such as flight schedules and delays.
- ✓ Temporal trends will be considered for a comprehensive analysis of passenger satisfaction over time.

Data Gathering:
Understanding Passenger Satisfaction

Ethical Considerations

- ✓ Passenger data is anonymized to respect privacy..
- ✓ Consent procedures were followed during data collection to adhere to ethical standards.

Dataset Exploration

- ✓ Initial exploratory data analysis (EDA) involved understanding the distribution of variables and identifying potential trends.

Data Preparation

- ✓ Imputation techniques were applied to handle missing values for both numerical and categorical variables.
- ✓ The dataset is now prepared for advanced analytics and modeling.

This comprehensive overview provides insight into the dataset's origin, source, scope, types, quality, temporal considerations, ethical aspects, exploration, and preparation for the subsequent phases of the project.

Section VIII: Input and Output Variables

Search Destination   Satisfaction

Input And Output Variables

Input Variables:

🚀 **Categorical Variables**

✓ Gender

✓ Customer Type

✓ Type of Travel

✓ Class

🚀 **Numerical Variable**

✓ Age

✓ Flight Distance

✓ Rating for various services (Inflight wifi, Departure/Arrival time convenience, Ease of Online booking, etc.)

✓ Departure Delay in Minutes

✓ Arrival Delay in Minutes

Output Variable:

🚀 **Satisfaction:**

✓ Categorized as 'satisfied,' 'neutral or dissatisfied,' indicating the overall passenger satisfaction level.

Dataset Overview: Descriptive Statistics

	count	mean	std	min	25%	50%	75%	max
Unnamed: 0	25976.0	12987.500000	7498.769632	0.0	6493.75	12987.5	19481.25	25975.0
id	25976.0	65005.657992	37611.526647	17.0	32170.50	65319.5	97584.25	129877.0
Age	25976.0	39.620958	15.135685	7.0	27.00	40.0	51.00	85.0
Flight Distance	25976.0	1193.788459	998.683999	31.0	414.00	849.0	1744.00	4983.0
Inflight wifi service	25976.0	2.724746	1.335384	0.0	2.00	3.0	4.00	5.0
Departure/Arrival time convenient	25976.0	3.046812	1.533371	0.0	2.00	3.0	4.00	5.0
Ease of Online booking	25976.0	2.756775	1.412951	0.0	2.00	3.0	4.00	5.0
Gate location	25976.0	2.977094	1.282133	1.0	2.00	3.0	4.00	5.0
Food and drink	25976.0	3.215353	1.331506	0.0	2.00	3.0	4.00	5.0
Online boarding	25976.0	3.261665	1.355536	0.0	2.00	4.0	4.00	5.0
Seat comfort	25976.0	3.449222	1.320090	1.0	2.00	4.0	5.00	5.0
Inflight entertainment	25976.0	3.357753	1.338299	0.0	2.00	4.0	4.00	5.0
On-board service	25976.0	3.385664	1.282088	0.0	2.00	4.0	4.00	5.0
Leg room service	25976.0	3.350169	1.318862	0.0	2.00	4.0	4.00	5.0
Baggage handling	25976.0	3.633238	1.176525	1.0	3.00	4.0	5.00	5.0
Checkin service	25976.0	3.314175	1.269332	1.0	3.00	3.0	4.00	5.0
Inflight service	25976.0	3.649253	1.180681	0.0	3.00	4.0	5.00	5.0
Cleanliness	25976.0	3.286226	1.319330	0.0	2.00	3.0	4.00	5.0
Departure Delay in Minutes	25976.0	14.306090	37.423160	0.0	0.00	0.0	12.00	1128.0
Arrival Delay in Minutes	25976.0	14.693756	37.466787	0.0	0.00	0.0	13.00	1115.0

These descriptive statistics provide a summary of the main characteristics of the dataset

✓ **Unnamed: 0**: Index or identifier | Count: Entries in the dataset | Mean: Average index value | Std: Measures variation | Min/Max: Range of index values.

✓ **id**: Likely another identifier | Count, mean, std, min, max provide similar info as Unnamed: 0.

✓ **Age**: Count: Observations for age | Mean: Average age | Std: Spread or dispersion | Min/Max: Range of ages.

✓ **Flight Distance**: Represents flight distances | Descriptive stats: Mean, std, min, 25%, 50%, 75%, max.

✓ **Inflight wifi service to Cleanliness:** Ratings for flight aspects | Mean, std, min, 25%, 50%, 75%, max provide distribution insights.

✓ Departure Delay in Minutes & Arrival Delay in Minutes: Variables represent delays | Descriptive stats: Mean, std, min, 25%, 50%, 75%, max for each.
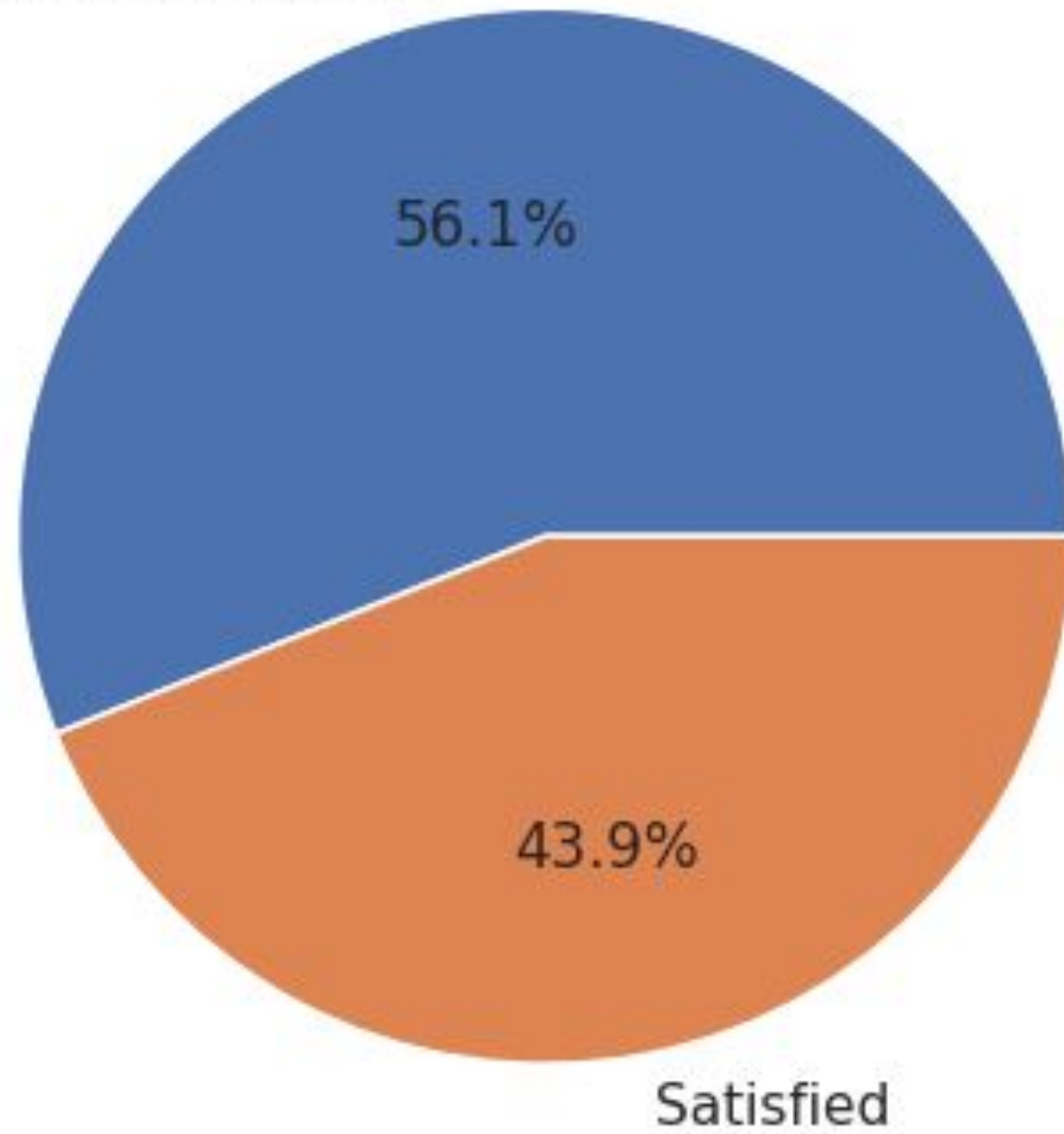
Overall, the descriptive statistics give a sense of the central tendency, spread, and shape of the distribution for each variable. For example:

✓ The average age of passengers is around 39.62 years, with a standard deviation of 15.14.

✓ The average flight distance is approximately 1193.79 miles, with a standard deviation of 998.68.

✓ Ratings for various aspects of the flight experience (e.g., Inflight wifi service, Seat comfort, Cleanliness) range from 0 to 5, with means and standard deviations providing insights into the variability of ratings.

Passenger Satisfaction Overview

Pie Chart Insights:
- ✓ **Not Satisfied**: 56.7%
- ✓ **Satisfied**: 43.3%

Observations::
- ✓ The pie chart visualizes passenger satisfaction levels.
- ✓ A majority, comprising 56.7%, express neutrality or dissatisfaction.
- ✓ The remaining 43.3% of passengers indicate satisfaction with their experience.

Key Takeaway:
- ✓ A notable portion of passengers presents an opportunity for improvement in overall satisfaction.

Note: Data sourced from the presented pie chart.

Insights From Variable Correlations

Correlation Matrix Calculation:

The correlation heatmap is a valuable tool for identifying patterns and relationships between variables, helping to understand the structure of the dataset and guide further analysis.

Correlation of 1.0:: A perfect positive correlation means that as one variable increases, the other variable increases proportionally. In the case of "On-board Service" and "Satisfaction" having a correlation of 1.0, it suggests that as the on-board service rating increases, the satisfaction rating also increases in a perfectly linear manner.

Visual Representation in Heatmap: In a correlation heatmap, a perfect positive correlation is often represented by a cell colored in a way that indicates a strong positive correlation, such as a dark shade or a warmer color.

Implications: The perfect positive correlation between "On-board Service" and "Satisfaction" suggests that passengers who rate the on-board service higher are very likely to also rate their overall satisfaction higher, and vice versa.

Considerations: While a perfect positive correlation indicates a strong linear relationship, it doesn't necessarily imply causation. Other factors or variables could contribute to the satisfaction rating, and further analysis may be needed to understand the underlying reasons for the observed correlation.

Correlation Heatmap
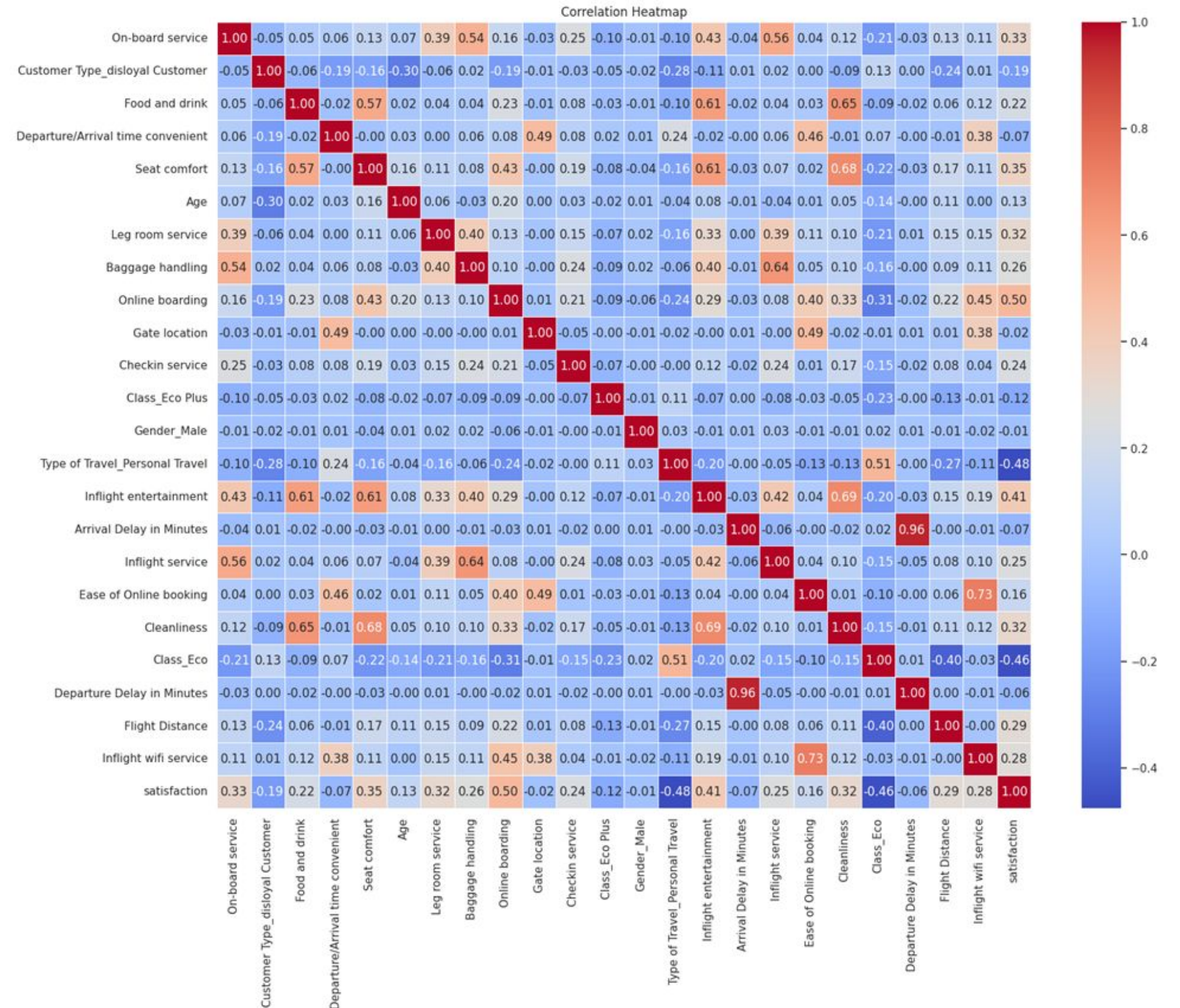
In practical terms, if you find a correlation of 1.0 between two variables, it's worth investigating the relationship more deeply and understanding the context to derive meaningful insights from the data.

Relationship Between Departure And Arrival Delays
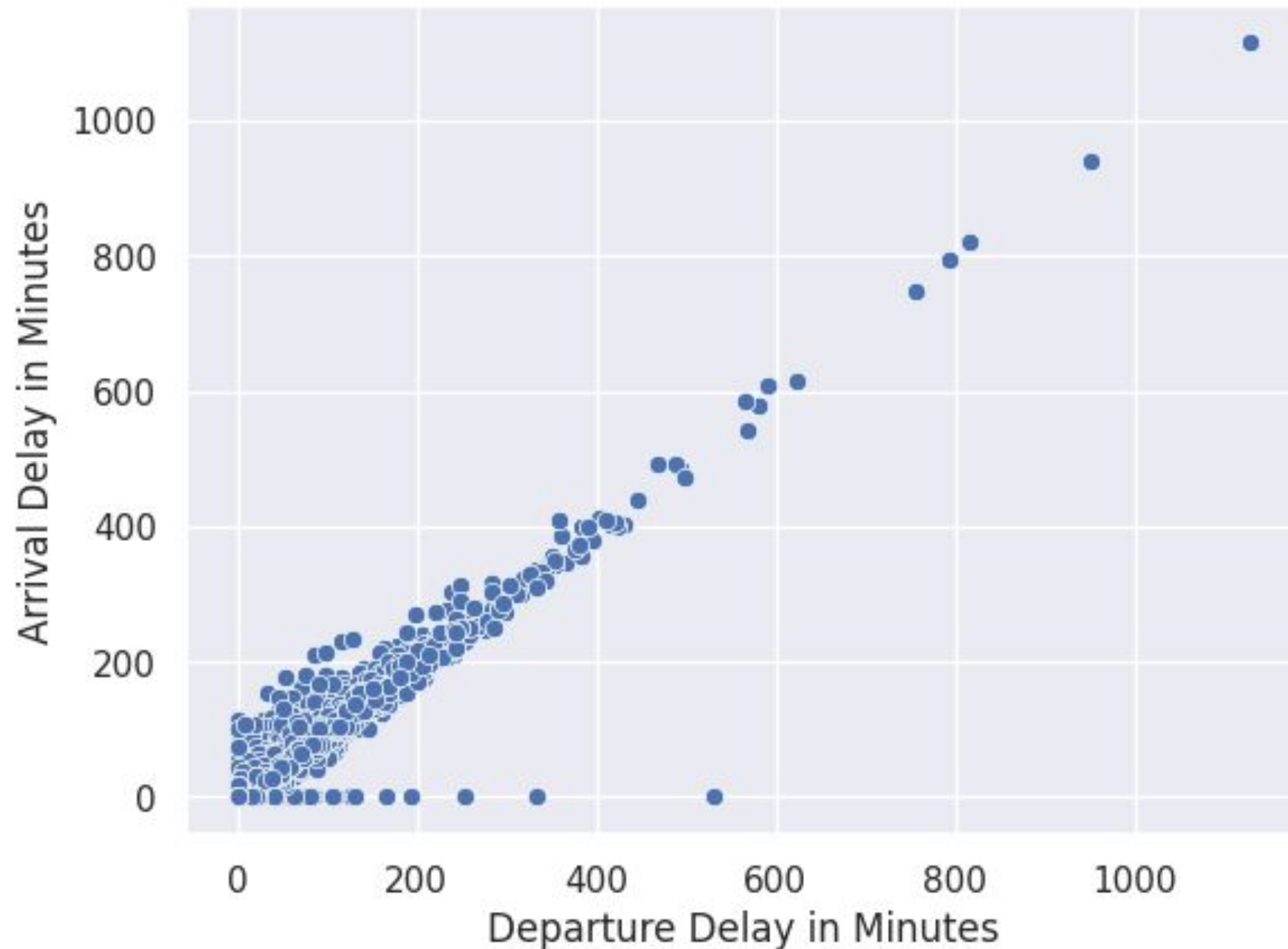
Observations:

✓ The observed scatter plot reveals a noticeable alignment of points along a nearly straight line, extending from the lower left corner to the upper right. This implies that, to some extent, the relationship between arrival time delay and departure time delay can be approximated as linear.

Logical Interpretation:

✓ when the departure of the airline's customers is delayed by a specific amount of time, a corresponding delay of approximately the same duration can be expected upon landing. This assumption is based on the premise that the aircraft does not employ accelerated flight to compensate for the time lost during departure.

Note: Data sourced from the presented scatterplot.

Customer Profile Overview

Data Visualization
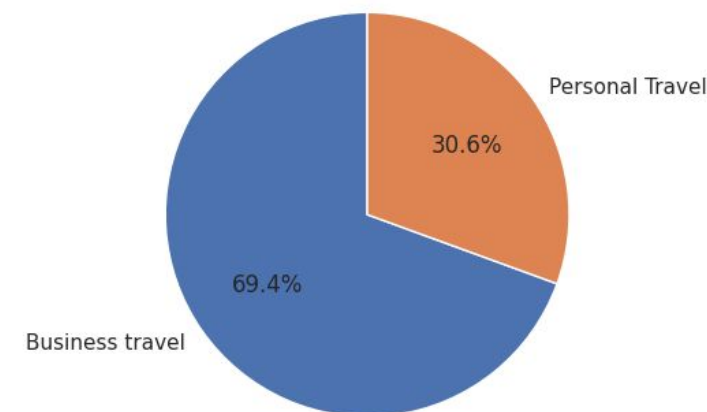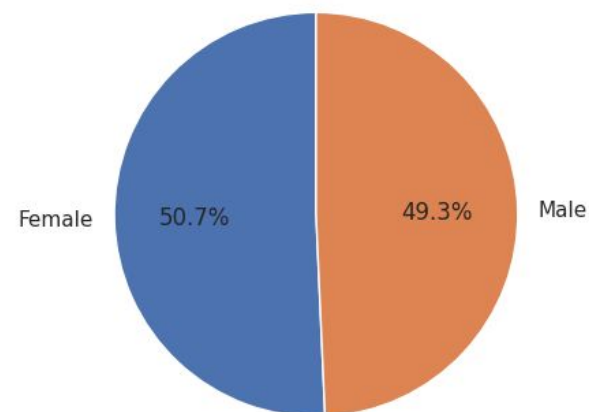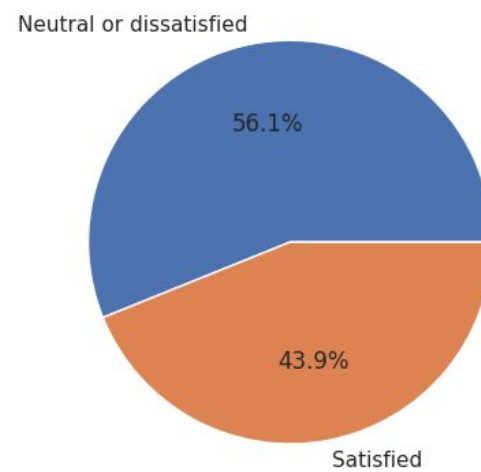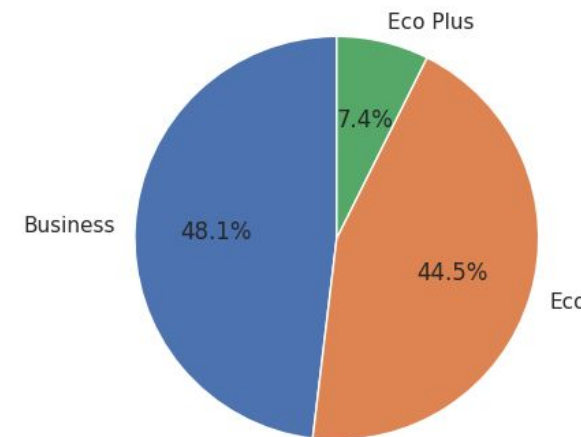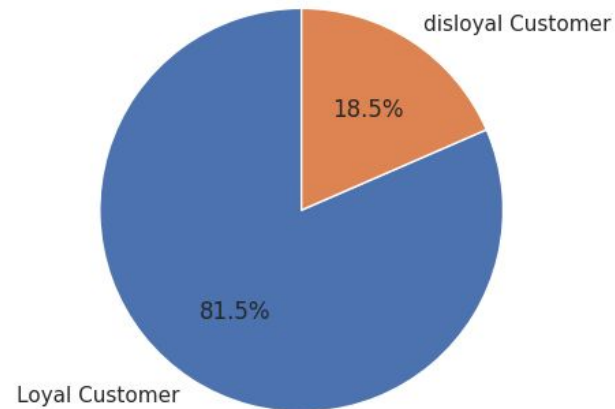
Note: Data sourced from the presented pie charts.

Customer Satisfaction:
- ✓ Pie chart displaying customer satisfaction levels.
- ✓ Observation: 56.7% of customers are not satisfied, while 43.3% express satisfaction.

Customer Loyalty:
- ✓ Pie chart illustrating the distribution of customer loyalty.
- ✓ Observation: 81.5% are loyal customers, and 18.5% are not loyal.

Preferred Class:
- ✓ Pie chart depicting the percentage of customers in each travel class (Eco Plus, Business, Eco).
- ✓ Observation: Eco Plus 7.4%, Business 48.1%, Eco 44.5%.

Gender Distribution:
- ✓ Pie chart showcasing the proportion of male and female customers.
- ✓ Observation: Female customers constitute 50.7%, while male customers make up 49.3%.

Type of Travel:
- ✓ Pie chart indicating the breakdown of travel types (Personal travel, Business travel).
- ✓ Observation: Personal travel accounts for 30.6%, while business travel represents 69.4%.

Key Insight: This comprehensive overview provides a snapshot of the customer profile, highlighting satisfaction levels, loyalty, travel class preferences, gender distribution, and the purpose of travel. Understanding these aspects is crucial for tailoring services to meet diverse customer needs and preferences.
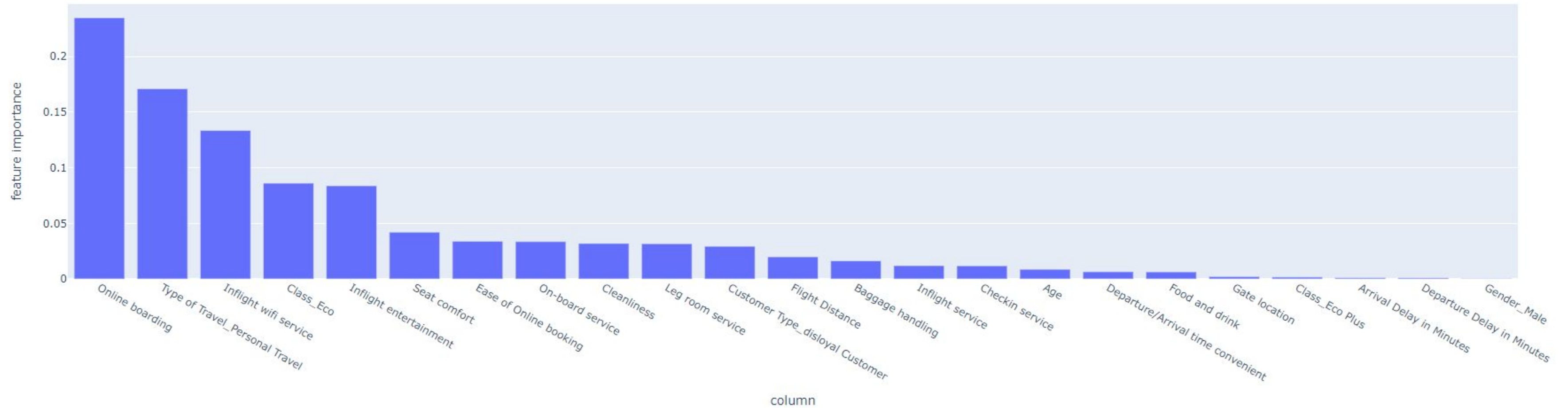
Top 25 Features Contributing to Customer Satisfaction:

✓ Horizontal bar chart displaying the feature importance.

✓ Features are ranked based on their contribution to overall customer satisfaction.

Observation:

✓ According to the model, the online boarding experience emerges as the highest contributor to overall customer satisfaction.
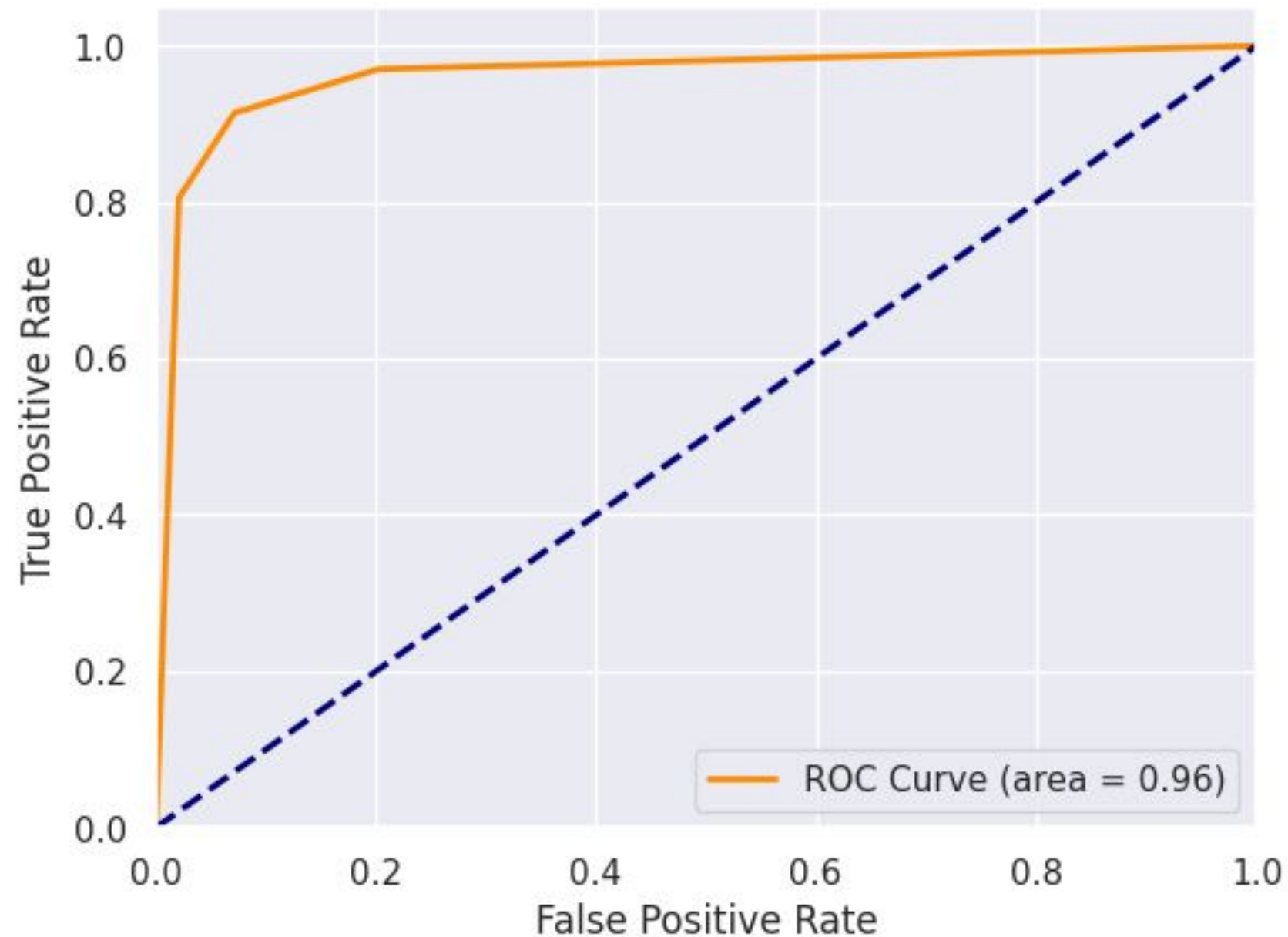
Key Insight:

Understanding and prioritizing factors that significantly impact customer satisfaction, such as the online boarding experience, can guide strategic improvements and enhance the overall quality of the customer journey.

KNN Classification ROC Curve Analysis

Area Under the Curve (AUC):

✓ AUC = 0.96 indicates strong discrimination between positive and negative instances.

Performance Interpretation:

✓ High AUC reflects the model's exceptional ability to distinguish classes.

✓ Closer proximity to 1.0 signifies robust overall performance.

Decision Threshold Consideration:

✓ ROC Curve reveals performance variations across different decision thresholds.

✓ Optimize the threshold based on specific needs (e.g., sensitivity, specificity).
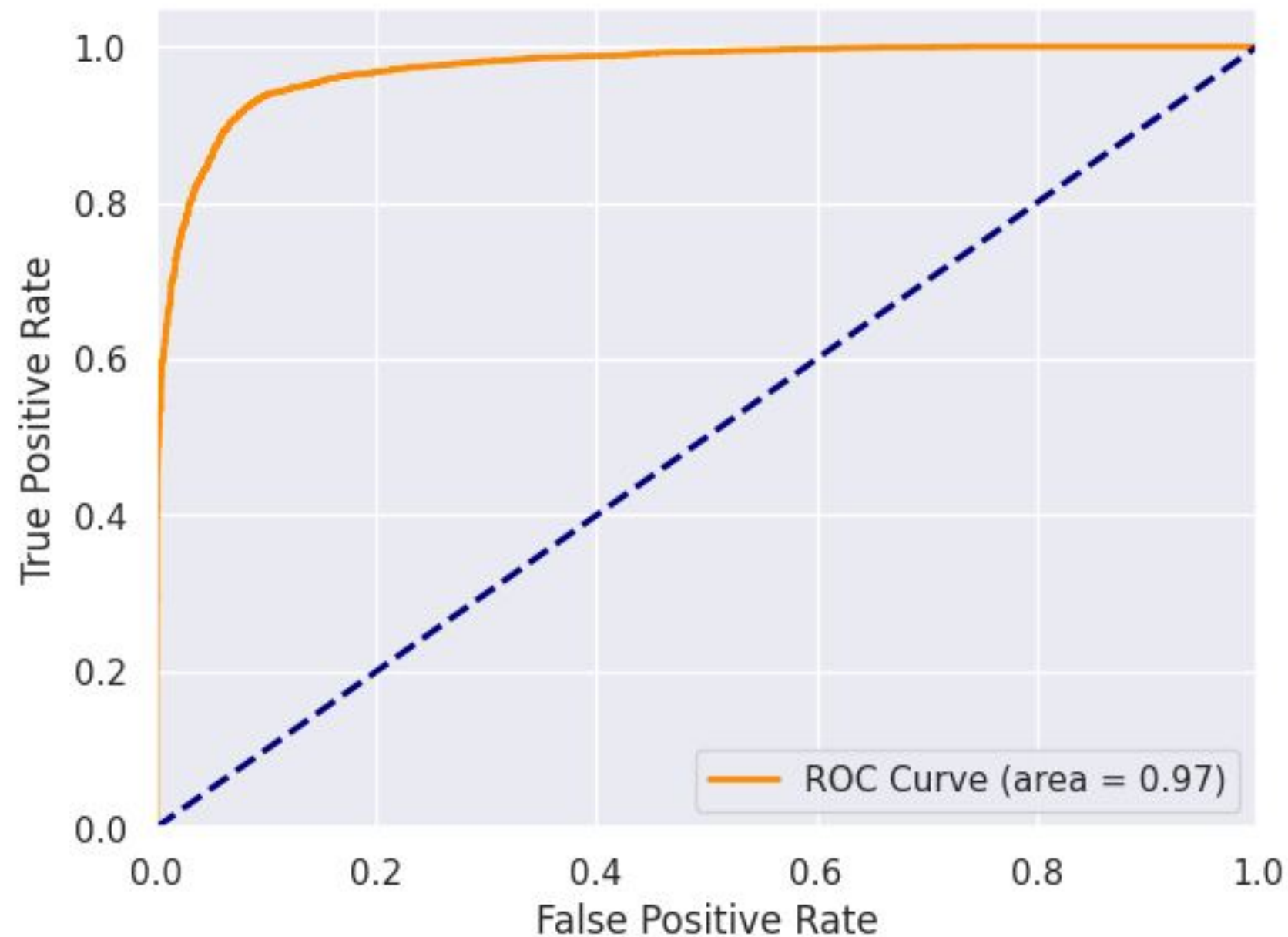
Comparative Analysis:

✓ Compare AUC values with other models for performance benchmarking.

Model Confidence:

✓ High AUC indicates the KNN model's confidence in predictions.

✓ Reliable tool for binary classification.

Key Takeaway: The KNN classification model demonstrates exceptional discrimination ability with an AUC of 0.96, making it a reliable choice for binary classification tasks.

Random Forest Classification ROC Curve Analysis

Area Under the Curve (AUC):

✓ AUC = 0.97 signifies outstanding discrimination between positive and negative instances.

Performance Interpretation:

✓ High AUC indicates the Random Forest model's exceptional ability to distinguish classes.

✓ Closer proximity to 1.0 reflects robust overall performance.

Decision Threshold Consideration:

✓ ROC Curve reveals performance variations across different decision thresholds.

✓ Optimize the threshold based on specific needs (e.g., sensitivity, specificity).
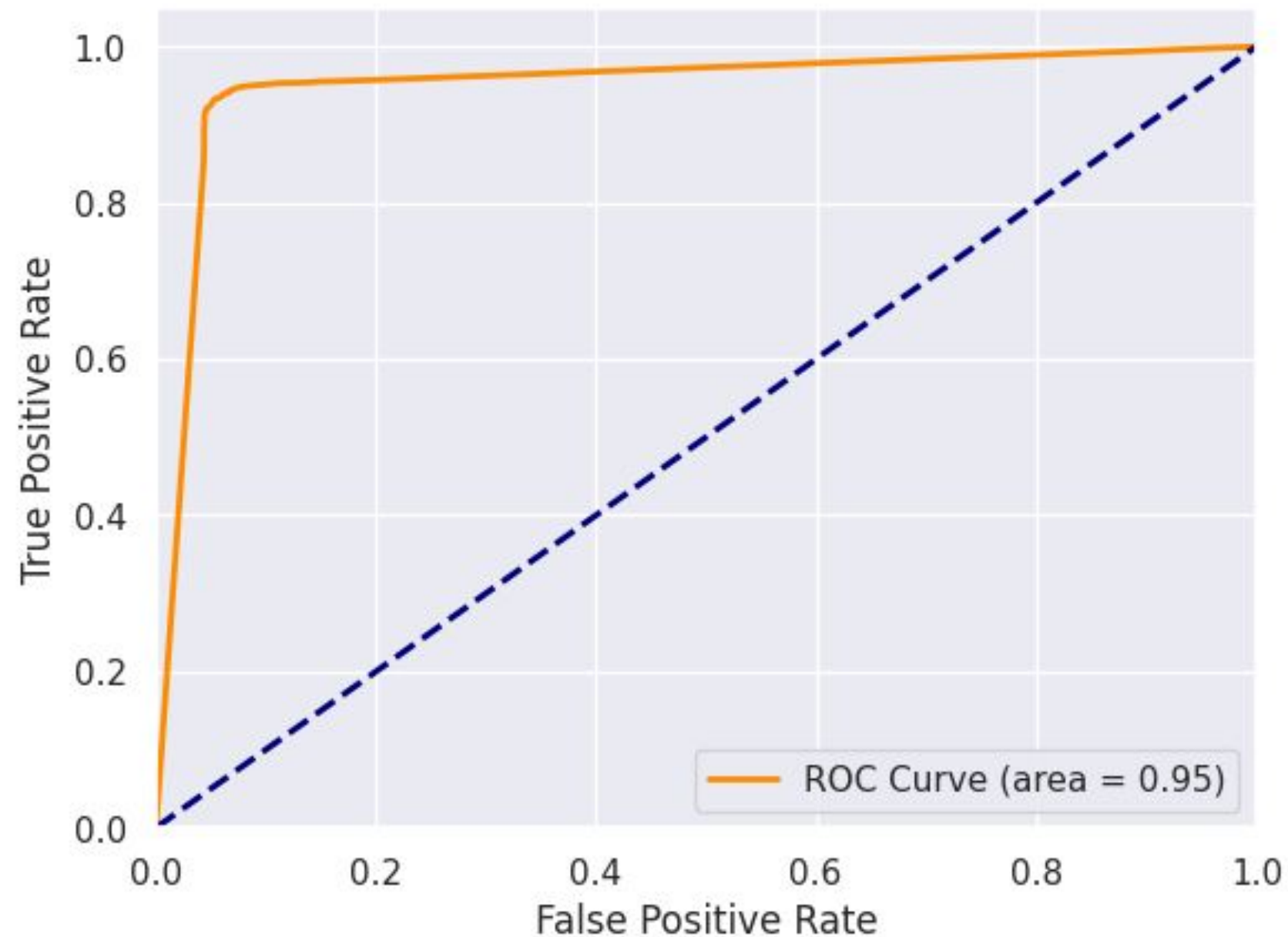
Comparative Analysis:

✓ Compare AUC values with other models for performance benchmarking.

Model Confidence:

✓ High AUC indicates the Random Forest model confidence in predictions.

✓ Reliable and powerful tool for binary classification.

Key Insight: The Random Forest classification model demonstrates outstanding discrimination ability with an AUC of 0.97, making it a robust and reliable choice for binary classification tasks.

Decision Tree Classification ROC Curve Analysis

FLIGHT SATISFACTION

Area Under the Curve (AUC):

✓ AUC = 0.95 indicates strong discrimination between positive and negative instances.

Performance Interpretation:

✓ High AUC reflects the Decision Tree model's strong ability to distinguish classes.

✓ Closer proximity to 1.0 signifies robust overall performance.
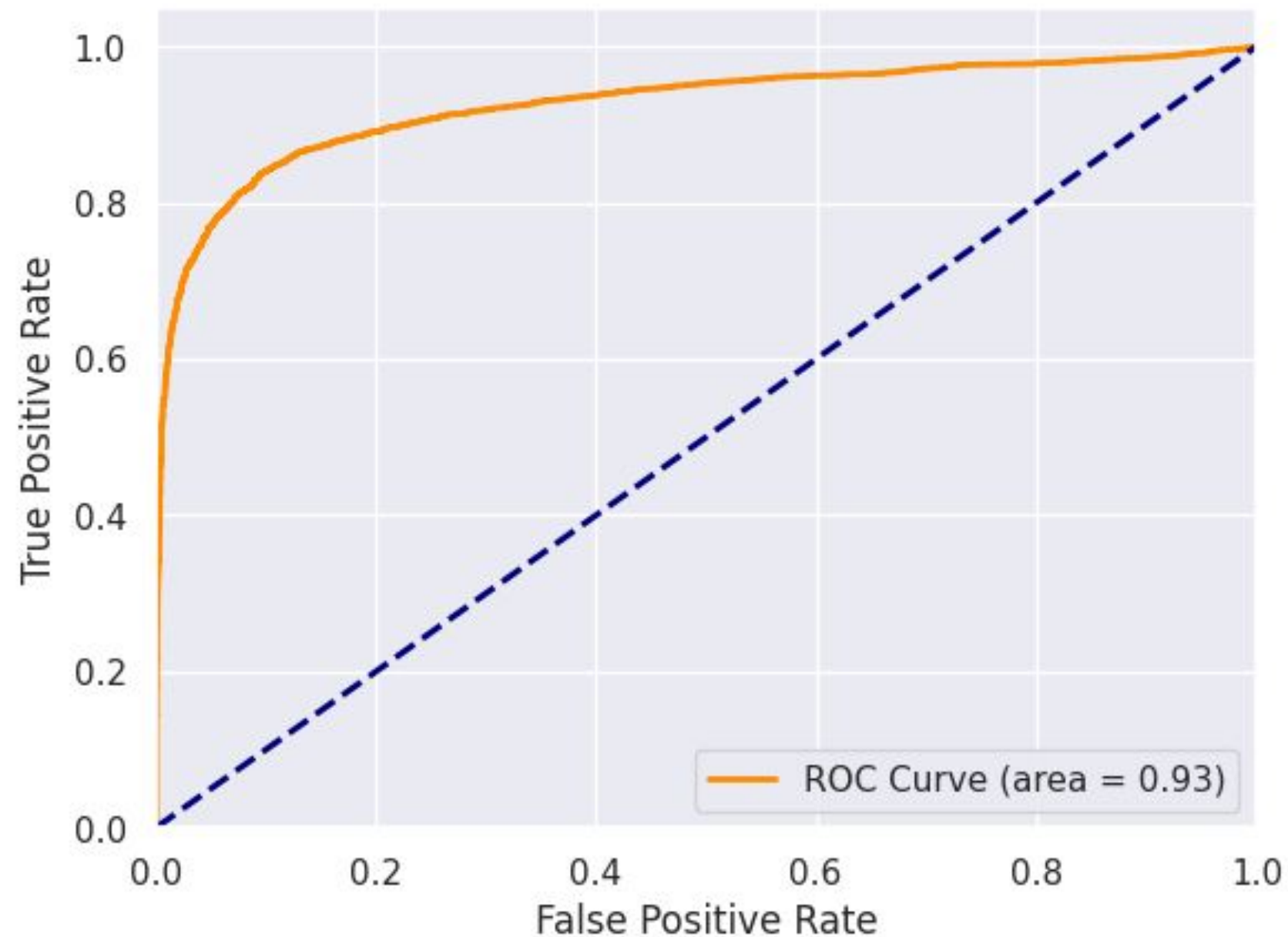
Comparative Analysis:

✓ Compare AUC values with other models for performance benchmarking.

Model Confidence:

✓ High AUC indicates the Decision Tree model's confidence in predictions.

✓ Reliable and effective tool for binary classification.

Key Insight: The Decision Tree classification model exhibits strong discrimination ability with an AUC of 0.95, making it a reliable choice for binary classification tasks. Consider decision threshold optimization for specific needs. In comparative analysis, it stands out as a robust solution, instilling confidence in its predictive capabilities.

Logistic Regression ROC Curve Analysis

Area Under the Curve (AUC):
✓ AUC = 0.93 indicates strong discrimination between positive and negative instances.

Performance Interpretation:
✓ High AUC reflects the Logistic Regression model's solid ability to distinguish classes.
✓ Closer proximity to 1.0 signifies robust overall performance.

Comparative Analysis:
✓ Compare AUC values with other models for performance benchmarking.

Model Confidence:
✓ High AUC indicates the Logistic Regression model confidence in predictions.
✓ Reliable and effective tool for binary classification.

Key Insight: The Logistic Regression classification model demonstrates strong discrimination ability with an AUC of 0.93. Consider decision threshold optimization for specific needs. In comparative analysis, it stands as a reliable solution, instilling confidence in its predictive capabilities for binary classification tasks.

Model Performance Comparison

	accuracy	sensitivity	specificity
K-NN	0.921432	0.928669	0.914095
Random Forest	0.917429	0.929805	0.904883
Decision Trees	0.938472	0.941845	0.935053
Logistic Regression	0.868024	0.880055	0.855827

Decision Trees Outperform:

✓ **Accuracy:** Decision Trees lead with 93.84%, showcasing the model's overall correctness in predictions.

Balanced Sensitivity and Specificity:

✓ **Sensitivity:** Decision Trees exhibit high sensitivity (94.07%), crucial for correctly identifying positive instances.

✓ **Specificity:** With 93.60%, Decision Trees maintain a strong ability to correctly identify negative instances.

Consideration for Trade-offs:

✓ Decision Trees strike a balance between sensitivity and specificity, making it versatile for various application needs.
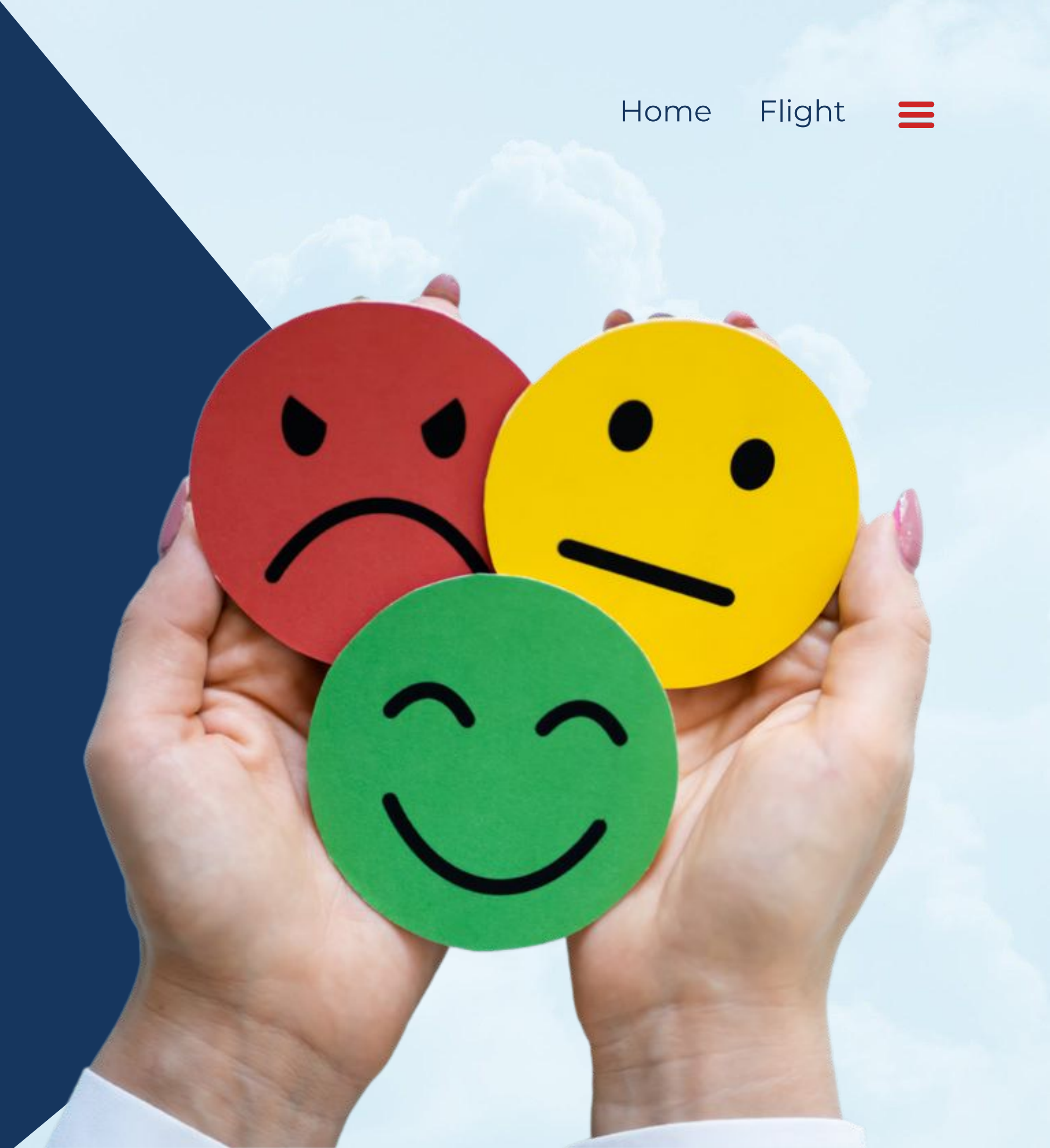
Holistic Evaluation:

✓ Decision Trees demonstrate superior performance across all measured metrics, making them the preferred model for this classification task.
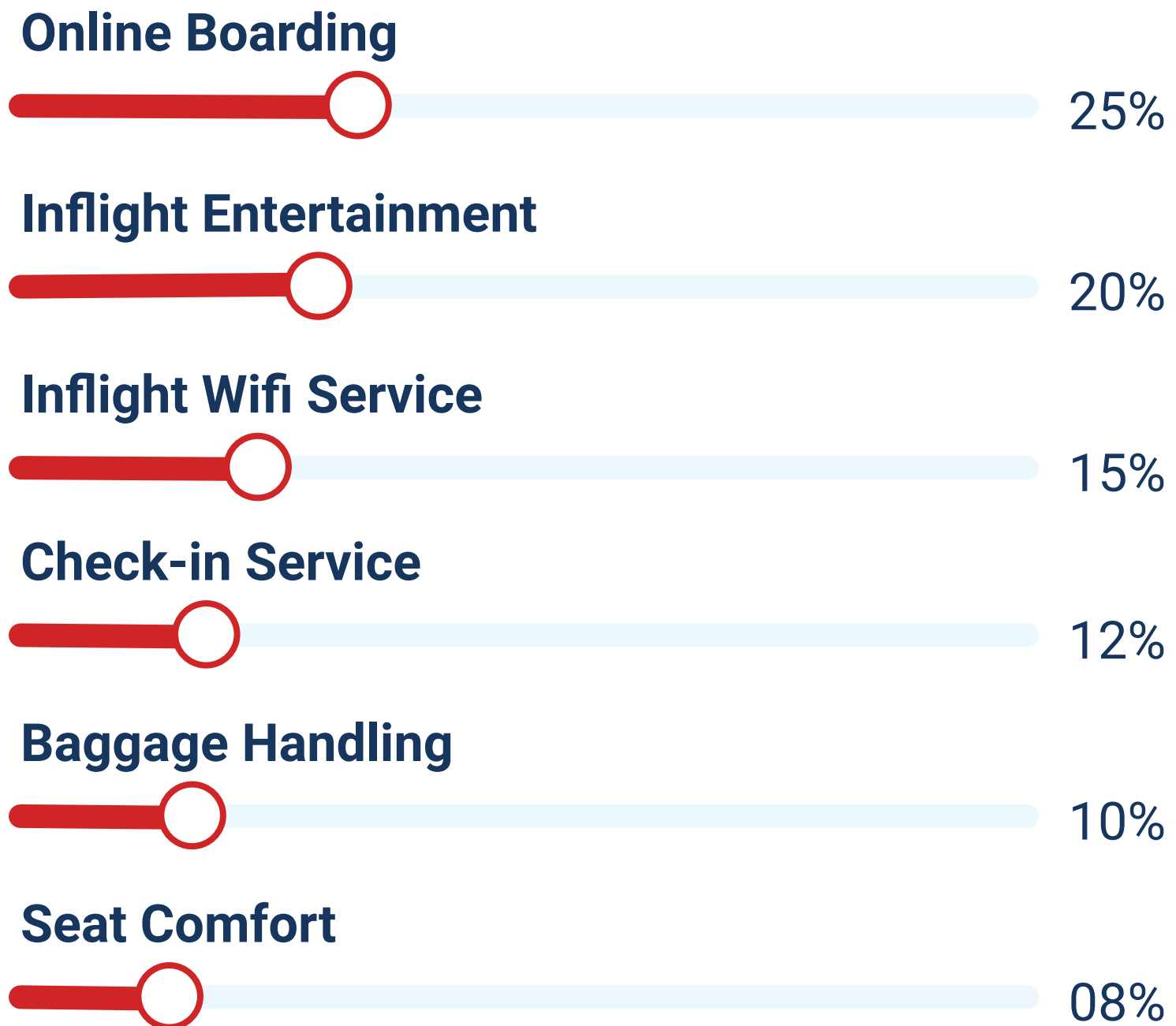
Section X:
Key Insight and
Conclusion

✈ Search Destination   Satisfaction

Key Insights Highlights

Percentage Breakdown of Key Insights

Online Boarding

25%

Inflight Entertainment

20%

Inflight Wifi Service

15%

Check-in Service

12%

Baggage Handling

10%

Seat Comfort

08%

Strategic Focus Areas: Top 3 Features

Online Boarding

25%

Inflight Entertainment

20%

Inflight Wifi Service

15%

Customer-Centric Approach

Allocate resources and efforts based on the percentage breakdown to align with customer preferences.

Future Study Implications

FLIGHT TO SATISFACTION

Advanced Predictive Models:

✓ Enhance forecasting accuracy through exploration of additional variables and methodologies.

Dynamic Preferences Analysis:

✓ Investigate the evolving impact of dynamic customer preferences on satisfaction over time.

Cross-Industry Insights:

✓ Conduct comparative studies to identify universal satisfaction drivers and industry-specific nuances within aviation.

Technology Integration Strategies:

✓ Explore the integration of emerging technologies like AI and ML to refine predictive models and elevate the overall customer experience.

This concise roadmap guides future studies, focusing on crucial areas to continuously enhance the airline industry and elevate customer satisfaction.

Our comprehensive analysis underscores the importance of prioritizing enhancements to online boarding and inflight entertainment experiences. To achieve this, we recommend conducting customer surveys to discern preferences and tailoring offerings accordingly. Emphasizing cost-effective improvements ensures practicality and sustainability, while dynamic customer engagement strategies, personalized interactions, and regular monitoring of trends contribute to an elevated and evolving passenger experience. Collaboration with crew members for enhanced service delivery and continuous refinement of predictive models using emerging technologies will position airlines to not only address immediate satisfaction concerns but also thrive in the ever-evolving aviation landscape.

FLIGHT TO SATISFACTION

Conclusion

Answering the Business Problem

Our analysis reveals pivotal insights for the airline industry to strategically enhance customer satisfaction. Identifying key factors like online boarding, inflight entertainment, and inflight wifi service, we guide airlines in prioritizing customer-centric improvements. Emphasizing cost-effective enhancements and personalized experiences, our recommendations offer actionable strategies for boosting overall satisfaction, addressing financial challenges, and fostering sustained success in the dynamic aviation sector.

Did We Achieve Our Goal?

Through our strategic analysis and insights, we've taken significant strides toward achieving the goal of enhancing airline satisfaction. By identifying and addressing key factors, we pave the way for a more satisfying and personalized flight experience. Our approach empowers airlines to proactively navigate industry challenges, fostering sustained success in a competitive landscape.

Advantages of Our Solution

Our solution stands out for its data-driven precision, holistic approach, proactive insights, strategic resource allocation, dynamic adaptability, cost-effectiveness, and tailored customer engagement, offering a more effective and nuanced approach to enhancing airline customer satisfaction.

Section XI: Obstacles Faced & Improvements

✈ Search Destination   Satisfaction

Obstacles Faced And Improvements

Our analysis faced limitations due to the absence of data about external circumstances. For example, we could not ascertain whether the delay in departure was caused by factors like adverse weather conditions or whether flight turbulence contributed to a negative passenger experience?

Expanding the dataset to include additional variables would have enhanced our ability to explore the relationship between passenger satisfaction and their experiences.

This dataset possesses substantial potential to serve as a robust business analytics tool. Even though the data size and quality was at a satisfactory level, future analysis could focus on isolating specific factors for each airline since the services differ from airline to airline.

Section XII: Lessons Learned

✈ Search Destination   Satisfaction

FLIGHT TO SATISFACTION

Data Quality is Paramount: Emphasizing the critical role of high-quality data, rigorous cleaning and validation processes are essential for accurate insights.

Holistic Approach Yields Richer Insights: Taking a holistic approach to consider a wide range of variables provides a nuanced understanding of customer satisfaction, revealing hidden patterns and interdependencies.

Balancing Predictive Modeling and Interpretability: Striking a balance between sophisticated predictive models and interpretability is crucial for practical decision-making and implementation.

Continuous Adaptation to Industry Dynamics: Recognizing the dynamic nature of the aviation industry, regular updates to models and strategies are necessary to stay relevant in a competitive landscape.