AIRLINE CUSTOMER SATISFACTION

Team 1: Hema, Darsini, Hussain, Halima

The George Washington University

DATS 6103: Introduction to Data Mining

Contents

[Introduction: 1](#_Toc185352785)

[About the dataset: 2](#_Toc185352786)

[Data Cleaning: 3](#_Toc185352787)

[EDAs: 3](#_Toc185352788)

[Most important reason for satisfaction 3](#_Toc185352789)

[Loyalty vs satisfaction of Customers 4](#_Toc185352790)

[Key distribution trends among the travelers 5](#_Toc185352791)

[The satisfaction and dissatisfaction rate for different Age groups 7](#_Toc185352792)

[Modelling: 7](#_Toc185352793)

[Logistic Regression: 7](#_Toc185352794)

[Random Forest: 10](#_Toc185352795)

[Conclusions 10](#_Toc185352796)

[References 11](#_Toc185352797)

## Introduction:

Airline customer satisfaction is a critical factor in ensuring customer loyalty, driving revenue, and maintaining a competitive edge in the aviation industry. With increasing globalization and mobility, understanding the factors influencing customer satisfaction is vital for airlines to improve their services and retain passengers. This project aims to explore key factors influencing airline customer satisfaction and predict satisfaction rates using various service attributes, such as inflight WiFi, seat comfort, cleanliness, and others.

As international students living in the United States, we often think about flying to visit famous cities like New York, Los Angeles, Miami, Chicago, Denver, and many others. Domestic air travel plays a significant role in helping us explore the country and connect with friends, family, and new experiences. However, we’ve also noticed the challenges and expectations that come with flying on U.S. domestic airlines.

Through our experiences, we came to realize how important the aviation industry is to the United States. It not only supports millions of jobs but also drives economic growth and generates significant revenue. Customer satisfaction is crucial for ensuring a positive travel experience, as it directly influences passenger loyalty and the airline’s overall success. By analyzing and predicting satisfaction levels for U.S. domestic flights, we aim to provide actionable insights that can help airlines improve their services, enhance passenger experiences, and contribute to the continued strength and growth of the aviation industry.

Several studies have been conducted to analyze customer satisfaction in the airline industry, focusing on factors such as service quality, passenger expectations, and operational performance. To deepen our understanding of the topic, we reviewed multiple articles and reports. The J.D. Power North America Airline Satisfaction Study (2024) (JDPower, n.d.) identifies key factors influencing passenger satisfaction, such as cost, in-flight services, and staff performance, while highlighting ongoing challenges like crowded flights and delays. Additionally, research by Kandpal and Mehrotra (2020) (arxiv, n.d.) analyzed online passenger reviews, revealing that seat comfort, punctuality, and cleanliness are among the most significant concerns for travelers. These studies, along with other industry insights, have helped us better understand the critical role customer satisfaction plays in shaping the success of U.S. domestic airlines and how addressing these factors can lead to improved passenger experiences.

## About the dataset:

The dataset used in this project was obtained from Kaggle’s Airline Passenger Satisfaction dataset, accessible at [this link](https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction/data). The dataset contains over 100,000 records of airline passengers, providing detailed information about their demographics, flight details, service ratings, and satisfaction outcomes. This comprehensive dataset serves as the foundation for analyzing key factors influencing customer satisfaction and building predictive models.

The dataset includes the following columns:

* **Customer ID**: A unique identifier for each passenger.
* **Gender**: The gender of the passenger (Male/Female).
* **Age**: The age of the passenger.
* **Type of Travel**: Whether the travel was for business or personal reasons.
* **Class**: The flight class (Business, Economy, or Economy Plus).
* **Flight Distance**: The distance traveled during the flight.
* **Inflight WiFi Service**: Rating of the inflight WiFi service (1-5 scale).
* **Departure/Arrival Time Convenience**: Rating of the convenience of departure and arrival times (1-5 scale).
* **Ease of Online Booking**: Rating of the online booking process (1-5 scale).
* **Gate Location**: Rating of the gate location (1-5 scale).
* **Food and Drink**: Rating of food and beverage service (1-5 scale).
* **Online Boarding**: Rating of the online boarding experience (1-5 scale).
* **Seat Comfort**: Rating of seat comfort (1-5 scale).
* **Inflight Entertainment**: Rating of inflight entertainment options (1-5 scale).
* **On-board Service**: Rating of on-board service (1-5 scale).
* **Leg Room Service**: Rating of legroom (1-5 scale).
* **Baggage Handling**: Rating of baggage handling service (1-5 scale).
* **Check-in Service**: Rating of the check-in process (1-5 scale).
* **Inflight Service**: Overall rating of inflight service (1-5 scale).
* **Cleanliness**: Rating of cabin cleanliness (1-5 scale).
* **Satisfaction**: The target variable, indicating whether the passenger was satisfied (Satisfied/Not Satisfied).

## Data Cleaning:

Data cleaning was conducted to ensure reliable analysis and modeling by addressing missing values, encoding categorical features, and removing outliers. Missing values, particularly in "Arrival Delay in Minutes," were removed, while categorical variables like "Gender" and "Type of Travel" were label-encoded for compatibility with machine learning. As seen in below figures, Outliers in features such as "Flight Distance" and "Check-in Service" were identified using IQR and eliminated to prevent skewed results.

Before outlier removal: After outlier removal:

## EDAs:

### Most important reason for satisfaction

The violin plots illustrate the distribution of ratings for various flight services among satisfied and dissatisfied customers, shedding light on the key factors influencing customer satisfaction.

The analysis reveals that inflight WiFi service significantly impacts customer satisfaction, with satisfied customers consistently giving it higher ratings compared to dissatisfied customers. Other services, such as ease of online booking, seat comfort, inflight entertainment, and cleanliness, also show higher ratings among satisfied customers, indicating their importance in shaping a positive experience.

However, certain services like onboard service, legroom service, and gate location have ratings that are more similar between satisfied and dissatisfied customers. This suggests that while these factors contribute to the overall experience, they might not play as pivotal a role in distinguishing satisfaction levels.

The findings emphasize that improving critical services, such as inflight WiFi, seat comfort, and cleanliness, can significantly enhance customer satisfaction. Additionally, airlines can focus on optimizing onboard and legroom services to further improve the overall travel experience and retention rates.

Satisfied customers: Unsatisfied customers:

### Loyalty vs satisfaction of Customers

The graph below illustrates satisfaction levels across two customer segments, loyal customers (represented by 1) and non-loyal customers (represented by 0). This comparison provides insights into whether satisfaction rates vary between these groups.

The data reveals that satisfaction is significantly higher among loyal customers. The taller bar for loyal customers highlights a greater number of satisfied individuals in this group. On the other hand, the shorter bar for non-loyal customers reflects lower satisfaction levels within this segment.

This distribution is a positive sign for the business. Higher satisfaction rates among loyal customers align with the company's objective of prioritizing its most dedicated customer base. Importantly, the data suggests that non-loyal customers are not overwhelming the satisfaction metrics, which could have been a potential concern.

The graph underscores a clear trend, loyal customers are substantially more satisfied than their non-loyal counterparts. This insight helps the business better understand its audience and refine strategies to maintain and enhance satisfaction levels, particularly among its most valued customers. This outcome is a promising indicator of strong customer relationships and long-term growth potential.

### Key distribution trends among the travelers

Type of Travel: The analysis of travel types reveals that a significant majority of travelers, 68.2%, are business travelers. This dominant proportion highlights that airlines are primarily catering to individuals traveling for professional reasons. The remaining 31.8% comprises personal or leisure travelers, a notably smaller share of the customer base. These figures emphasize the importance of tailoring services, such as onboard amenities and scheduling, to meet the needs of business travelers, as they represent the bulk of the clientele.

A blue circle with a number of numbers

Description automatically generated

1. Class: When examining the distribution of travel classes, Business and Economy stand out as the predominant choices among passengers, with Business class accounting for 48.9% and Economy class closely following at 44.1%. Economy Plus, however, represents only 6.9% of travelers, making it a less favored option. This trend suggests that while a substantial number of passengers value the comfort and perks of Business class, a nearly equal share prioritizes the affordability of Economy class. The minimal representation of Economy Plus indicates either limited availability or low perceived value, warranting further exploration to enhance its appeal.

A pie chart with numbers and a number of percentages

Description automatically generated

1. Satisfaction: Customer satisfaction rates indicate that 54.6% of passengers are dissatisfied, whereas 45.4% report being satisfied. This distribution underscores a notable disparity, with dissatisfaction outweighing satisfaction among travelers. Such findings suggest room for improvement in addressing passenger concerns and expectations. Focusing on the key determinants of satisfaction, such as travel experience, punctuality, and onboard services, could help airlines bridge this gap and foster a more favorable perception among their customers.

A blue circle with a number of percentages

Description automatically generated

### The satisfaction and dissatisfaction rate for different Age groups

Using age categories, we observe clear trends:

A graph of different colored columns

Description automatically generated

* 18-30 Age Group: This group shows the highest dissatisfaction levels, suggesting that their expectations, possibly related to technology, comfort, or convenience, are not being adequately addressed.
* 41-50 Age Group: Passengers in this age group report the highest satisfaction, indicating that the services and experiences provided by the airline align well with their needs and preferences, such as stability, reliability, or value for money.
* This analysis highlights the importance of tailoring services to different age demographics to improve overall satisfaction. Airlines may need to explore targeted strategies for younger passengers, focusing on innovative solutions to enhance their travel experience, while maintaining high-quality offerings for their satisfied older demographic.

## Modelling:

### Logistic Regression:

The analysis aimed to identify the key factors influencing customer satisfaction in the airline industry and to build a predictive model. Initially, a Variance Inflation Factor (VIF) analysis was conducted to assess multicollinearity among features. From the image attached below, two features, **"Departure Delay in Minutes"** and **"Arrival Delay in Minutes"**, exhibited high VIF values, indicating a strong correlation. This result aligns with the expected relationship, as a delay in departure often leads to a delay in arrival. To address this redundancy, "Departure Delay in Minutes" was dropped, simplifying the model without losing significant predictive power.

A correlation matrix was then plotted to identify the most influential features on customer satisfaction (image attached below). Features with correlation coefficients greater than 0.5 were selected for model building, including **Class, Online Boarding,** and **Type of Travel**. These factors emerged as the primary drivers of satisfaction, while others, like inflight entertainment and WiFi service, did not meet the threshold for inclusion despite their potential relevance to the customer experience. This streamlined the focus of the model on the most impactful variables.

The dataset was split into training (80%) and testing (20%) subsets, ensuring a balanced representation of the target variable, **satisfaction**. A logistic regression model was developed using the selected features. Logistic regression is a statistical method used for binary classification. It predicts the probability of an outcome (e.g., satisfied or not satisfied) based on input features (IBM, n.d.). The model achieved an accuracy of approximately **83%**, demonstrating strong performance in predicting customer satisfaction. The confusion matrix attached below showed high true positive (TP) and true negative (TN) values, indicating the model's reliability in distinguishing between satisfied and dissatisfied customers. Summary of the model generated is attached below.

Model summary: Confusion matrix:

Further validation was performed using a Receiver Operating Characteristic (ROC) curve (Towards Datascience, n.d.). The curve revealed an **AUC score of 0.9**, signifying excellent model performance and a robust ability to discriminate between the two satisfaction classes. This result reinforces the model's effectiveness in capturing the underlying relationships between the selected features and customer satisfaction.

The analysis highlights that features such as **Class, Online Boarding,** and **Type of Travel** play a pivotal role in shaping customer satisfaction. Higher travel classes offer enhanced comfort and amenities, while the ease of online boarding streamlines the travel process. The type of travel also influences satisfaction, as business travelers may prioritize punctuality and inflight services differently than leisure travelers. By focusing on improving these services and minimizing delays, airlines can significantly enhance customer satisfaction and loyalty.

### Random Forest:

## Conclusions

The study identifies multiple factors as central to shaping customer satisfaction in the airline industry, with online boarding emerging as a critical determinant. The demand for a seamless and efficient check-in process has grown substantially among passengers, and online boarding directly addresses this need. By reducing wait times and offering unmatched convenience, this feature establishes a positive tone for the journey and contributes to elevated satisfaction metrics. Airlines that adopt and refine online boarding processes are likely to see significant improvements in customer loyalty and overall ratings.

Another pivotal factor is the availability of in-flight Wi-Fi, which has transitioned from being a supplementary amenity to a core expectation among passengers. For business travelers, uninterrupted connectivity is essential for productivity, while leisure passengers increasingly value the ability to access entertainment and stay connected during flights. Providing consistent and reliable in-flight Wi-Fi not only meets these diverse needs but also enhances the airline’s reputation and competitiveness within the market.

This type of travel influences customer satisfaction in nuanced ways. Business travelers, characterized by higher expectations for comfort and efficiency, represent a key demographic whose satisfaction carries significant weight. Their frequent travel patterns and willingness to pay for premium services underscore the need for tailored amenities. On the other hand, leisure travelers, while more cost-conscious, still appreciate thoughtful enhancements that improve their overall experience.

Additionally, the class of service directly correlates with satisfaction levels. Premium classes, such as Business or First Class, offer passengers enhanced comfort, superior dining, and personalized service, fostering consistently higher satisfaction scores. However, given that most passengers travel in Economy class, even modest improvements in comfort and service in this category can yield a meaningful impact on overall satisfaction ratings.

The analysis further highlights the superior performance of the Random Forest Classifier in identifying the primary drivers of customer satisfaction. Its ability to manage complex datasets and extract feature importance positions it as a valuable tool for data-driven decision-making. This finding underscores the importance of employing advanced analytical techniques to uncover actionable insights that can guide airline strategies.

In conclusion, the study reinforces the importance of adopting a passenger-centric approach to service delivery. Focusing on enhancements in online boarding, in-flight connectivity, class-specific offerings, and differentiated services for varying traveler demographics provides a robust framework for improving satisfaction. Airlines that prioritize these strategies are well-positioned to exceed customer expectations, foster loyalty, and maintain a competitive advantage in a demanding industry.

# References

(n.d.). Retrieved from IBM: https://www.ibm.com/think/topics/logistic-regression

(n.d.). Retrieved from Towards Datascience: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

(n.d.). Retrieved from arxiv: https://arxiv.org/abs/2012.08000

(n.d.). Retrieved from JDPower: https://www.jdpower.com/business/press-releases/2024-north-america-airline-satisfaction-study