Contents

[Introduction: 1](#_Toc185295924)

[About the dataset: 1](#_Toc185295925)

[Data Cleaning: 1](#_Toc185295926)

[EDAs: 2](#_Toc185295927)

[Most important reason for satisfaction 2](#_Toc185295928)

[Loyalty vs satisfaction of Customers 2](#_Toc185295929)

[Modelling: 3](#_Toc185295930)

[Conclusions 6](#_Toc185295931)

[References 6](#_Toc185295932)

## Introduction:

## About the dataset:

## 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.

## Modelling:

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.

## Conclusions

# 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