Abstract

This project aims to explore factors influencing airline passenger satisfaction using a dataset of over 4,000 customer reviews. We applied data cleaning, visualization, and logistic regression techniques. Key findings suggest that seat comfort, online boarding, and delay times are significant predictors of satisfaction. The project demonstrates how statistical modeling and visual analysis can support decision-making in the airline industry.

Introduction

Customer satisfaction plays a vital role in the airline industry, affecting customer retention, loyalty, and overall brand reputation. This analysis explores which aspects of service contribute most to a passenger's satisfaction. The motivation stems from increasing competition among airlines, making data-driven insights crucial for operational improvements.

Research Questions:

- What service factors most influence customer satisfaction?
- How do travel delays impact satisfaction?
- Can we predict satisfaction using logistic regression?

Data Description

The dataset contains **4,010 observations** with the following key variables:

Variable	Description
Satisfaction	Customer satisfaction (satisfied/dissatisfied)
Age	Passenger age
Class	Flight class (Eco, Eco Plus, Business)
Type of Travel	Purpose of travel (Business or Personal)
Seat comfort	Rating of seat comfort (1–5)
Food and drink	Rating of food quality (1–5)
Gate location	Rating of gate convenience
Inflight entertainment	Rating of entertainment options
Ease of Online booking	Ease of using the booking system
Baggage handling	Rating of baggage service
Departure/Arrival Delay	Minutes delayed
Flight Distance	Flight distance in miles

Methodology

We conducted descriptive statistical analysis to explore data distribution and group-wise comparisons. Missing values were handled using **mean imputation** for numeric columns. Key relationships were explored using visual tools such as bar plots, box plots, and histograms. Correlation analysis was conducted to examine multicollinearity. Finally, a **binary logistic regression** model was used to predict satisfaction.

Results

Key findings:

- Passengers in **Business class** were more satisfied than those in Economy.
- Higher ratings in **seat comfort**, **online boarding**, and **entertainment** were associated with satisfaction.
- Passengers experiencing **shorter delays** tended to report higher satisfaction.
- Logistic regression identified significant predictors including **seat comfort**, **online boarding**, and **flight distance**.

Discussion

These results align with expectations that **service quality** and **punctuality** are essential for customer satisfaction. However, limitations include potential bias in self-reported ratings and limited external variables (e.g., ticket price, staff behavior). Moreover, causality cannot be inferred. Still, the model provides valuable insights for airline management.

Conclusion

In summary, this project demonstrates the use of statistical analysis and predictive modeling to explore customer satisfaction in air travel. The findings highlight actionable areas for improvement, particularly in **inflight service** and **delay management**. Future work can incorporate time-series or **text-based sentiment analysis** using customer reviews for deeper insights.