

IDS 702 HW 5

Kayla Haeussler

Instructions: Use this template to complete your assignment. For this assignment, all code will be hidden in your rendered document. Therefore, you are required to submit both your qmd (this file) AND the rendered PDF. **Remember to render frequently**, as this will help you to catch errors in your code before the last minute.

Part 1: Data Science Team Report

Data Overview

This dataset pertains to a survey on customer flight experience and satisfaction, with each row representing an individual passenger. This dataset has 3,478 rows and 24 columns. The dataset does not contain any missing values.

Below is a summary of variables available in our dataset:

Variable	Median	Mean	Standard Dev
Age	40.0	39.67	15.11
Flight Distance	845.5	1186.76	984.99
Inflight Wifi Service	3.0	2.73	1.34
Ease of Online Booking	3.0	2.76	1.41
Gate Location	3.0	2.99	1.29
Food and Drink	3.0	3.22	1.33
Online Boarding	4.0	3.25	1.36
Seat Comfort	4.0	3.46	1.33
Inflight Entertainment	4.0	3.38	1.34
Onboard Service	4.0	3.41	1.28
Legroom Service	4.0	3.36	1.34
Baggage Handling	4.0	3.66	1.17
Check-in Service	3.0	3.32	1.26
Inflight Service	4.0	3.64	1.20
Cleanliness	3.0	3.29	1.32
Departure Delay (min)	0.0	14.50	38.81
Arrival Delay (min)	0.0	15.20	39.37

Analysis Plan

Model Type: An ordinal logistic regression model would be most appropriate in this case as customer satisfaction is our dependent variable, which is measured in an ordinal scale (dissatisfied, neutral, satisfied). We are interested in determine which factors influence customer satisfaction.

Link Function: The link function we will use in this case is the logit function

$$\text{logit}(P(Y \leq k)) = \log \left(\frac{P(Y \leq k)}{P(Y > k)} \right)$$

Predictors: Predictors for our model can be categorized into the following subsections

- Customer Demographic: Gender, age, and customer type (loyal vs. non-loyal).
- Travel Details: Type of travel (business vs. leisure), flight class (economy vs. business), and flight distance.
- Service Evaluations: Ratings for inflight experience (WiFi, seat comfort, etc.), ground services (check-in, booking), and cleanliness.
- Flight Timing: Departure and arrival delays.
-

Model Evaluation: We will evaluate our model based on its overall accuracy, using the No Information Rate, as well as the kappa statistic. Additionally we will use class-specific metrics like precision and sensitivity/recall and confusion matrix analysis.

Model Results

Predictor	Odds Ratio	Lower CI	Upper CI	P-Value
Gender (Male)	0.99	0.86	1.15	0.938
Customer Type (Loyal Customer)	5.24	4.17	6.60	<0.001
Age	0.99	0.99	1.00	0.034
Type of Travel (Personal)	0.14	0.11	0.17	<0.001
Class (Eco)	0.70	0.58	0.86	<0.001
Class (Eco Plus)	0.53	0.40	0.72	<0.001
Flight Distance	1.00	1.00	1.00	0.742
Inflight Wifi	1.37	1.25	1.50	<0.001
Service Departure / Arrival Time Convenience	0.95	0.90	1.013	0.127
Ease of Online Booking	0.75	0.68	0.82	<0.001
Gate Location	0.99	0.93	1.06	0.802
Food and Drink	0.97	0.90	1.06	0.589
Online boarding	1.51	1.40	1.63	<0.001
Seat comfort	1.06	0.97	1.15	0.191
Inflight Entertainment	1.06	0.95	1.74	0.299

Predictor	Odds Ratio	Lower CI	Upper CI	P-Value
On Board Service	1.17	1.09	1.25	<0.001
Leg Room Service	1.18	1.11	1.25	<0.001
Baggage Handling	1.08	0.99	1.73	0.069
Checkin Service	1.21	1.14	1.29	<0.001
Inflight Service	1.13	1.04	1.23	0.0029
Cleanliness	1.12	1.02	1.23	0.022
Departure Delay in Minutes	1.00	1.00	1.01	0.199
Arrival Delay in Minutes	0.99	0.99	1.00	0.0415

Key Results:

Based on the results of the ordinal logistic regression model, loyalty, travel purpose, class, inflight service and convenience appear to be the most influential predictors of customer satisfaction. Loyal customers are significantly more likely to report higher satisfaction levels, with odds 5.24 times greater than other customer types. Conversely, customers traveling for personal reasons are much less likely to be satisfied compared to business travelers. Class of service also plays a critical role, with passengers in Economy and Economy Plus showing lower odds of satisfaction compared to those in Business class. Among service-related factors, inflight Wi-Fi, online boarding, inflight service, leg room service, on-board service, check-in service, and cleanliness are strong positive predictors of satisfaction, indicating their importance in shaping customer experiences. Ease of online booking, on the other hand, is negatively associated with satisfaction, suggesting that challenges in this area may reduce customer contentment. Demographic factors like age show a slight decrease in satisfaction with older age groups, while gender has no significant effect. Flight distance and delays have minimal or no significant impact on satisfaction levels. Overall, loyalty, travel purpose, class of service, and various service-related factors emerge as the most influential predictors of customer satisfaction.

Model Assessment

Ordinal Logistic Regression

Prediction (below) / Actual (right)	dissatisfied	neutral	satisfied
dissatisfied	502	511	76
neutral	300	277	114

Prediction (below) / Actual (right)	dissatisfied	neutral	satisfied
satisfied	186	144	1368

Model Assessment Metrics:

- Accuracy: 61.73% - Higher than random chance, with relatively strong performance in predicting the “satisfied” category.

Kappa: 0.4012 - Reflects moderate agreement between predicted and actual satisfaction levels.

Balanced Accuracy: 56.73% - From this we see there is difficulties in accurately predicting less represented categories, such as “neutral” customers.

Multinomial Logistic Regression

Prediction (below) / Actual (right)	dissatisfied	neutral	satisfied
dissatisfied	464	423	116
neutral	341	368	61
satisfied	183	141	1381

Model Assessment Metrics:

- Accuracy: 63.63% - Slightly higher than ordinal logistic regression

Kappa: 0.4309 - Reflects slightly better overall classification than the ordinal model.

Balanced Accuracy: 61.85% - Slightly higher than the ordinal logistic regression

Model Assessment Conclusion

The Multinomial Logistic Regression performs better overall, with higher accuracy, Kappa, and balanced accuracy. This suggests it handles the data and class imbalances more effectively than the ordinal logistic regression model. However, the ordinal logistic model may still be preferable as we value interpretability more in this case.

Part 2: Executive Summary

Introduction

The purpose of our analysis was to investigate which aspects of a LaneAir customer's experience has the most impact on the overall satisfaction. Our analysis utilized a survey dataset which contains responses from 3,478 customers, including demographic details about each customer, flight details and service ratings, and a overall flying experienced rating of 'satisfied', 'neutral' or 'dissatisfied'. Of the customers survey, 52% said they were 'satisfied' with their flight experience, while 23% said 'neutral' and 23% responded 'dissatisfied'. Our respondents were 51% Female, 49% Male, with a median age of 40 years old. Average departure delay for respondents was 14.5 minutes, while average arrival delay was 15.2 minutes. Particularly low service ratings were those of inflight wifi and ease of online booking.

Methods

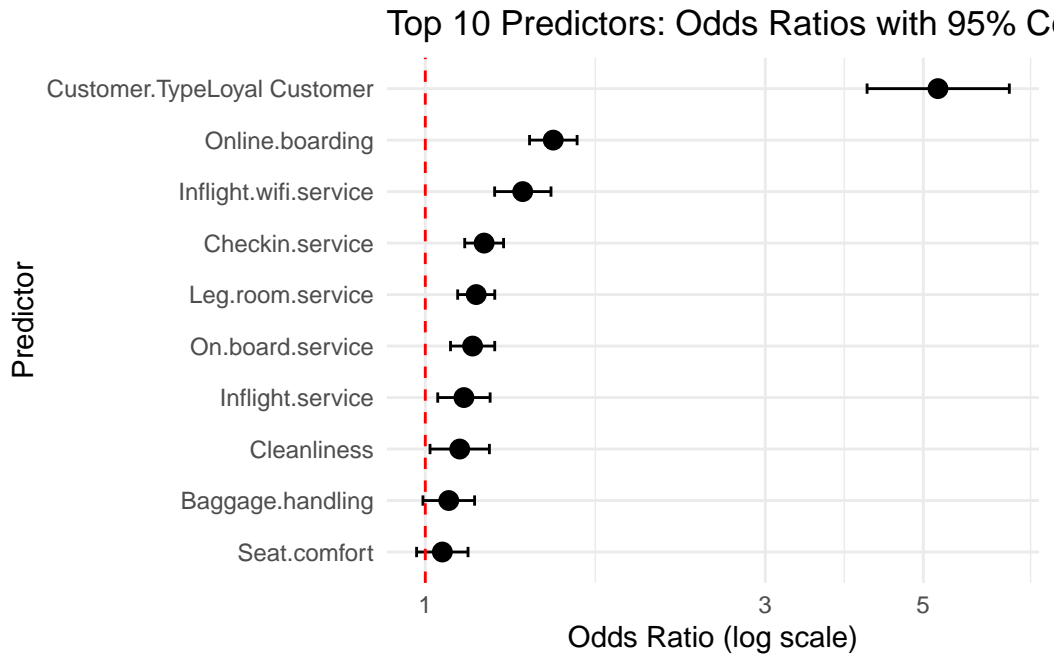
The analysis employs a combination of data preprocessing, statistical modeling, and visualizations to answer LaneAir's central question: "What factors most influence customer satisfaction?" Ordinal logistic regression was initially used due to its suitability for predicting ordinal categorical outcomes, which in this case was satisfaction levels. In addition to OLR, we also utilized multinomial logistic regression as it treats the satisfaction levels as individual categories, and identifies unique drivers at each of the satisfaction levels.

Results

From our analysis we found the following results:

- Loyal customers are 5 times more likely to be satisfied than disloyal customers
- Customers traveling for personal reasons as opposed to business or other reasons have a significantly less likely chance of reporting satisfaction.
- Customers who left higher ratings for online boarding are 1.5 times more likely to be satisfied
- Other aspects of service such as leg room service, inflight wifi, on board service and checkin service all show strong positive effects on satisfaction

Below is a key visualization highlighting customer satisfaction levels across different service attributes:



Conclusion

Based on the findings, the following recommendations are proposed:

1. **Service enhancement:** Focus on training staff to improve interactions with customers, particularly during check-in and in-flight services.
2. **Seats:** Increase legroom and space for customers on flights
3. **Wifi:** Ensure planes are equipped with high quality wifi services
4. **Customer Loyalty:** foster relationships with loyal customers to ensure they feel valued and continue to be repeat customer

However, given cost restraints, it is recommended that LaneAir invest resources in customer loyalty marketing initiatives and technology investments to improve wifi services, as our study suggested that both were important factors in influencing customer satisfaction, while also being relatively easier to implement for LaneAir

Study Limitations: The analysis is subject to potential response bias, as customers with extreme experiences (either very positive or negative) may be more likely to complete the survey. Furthermore, demographic or contextual variables not included in the dataset might influence satisfaction. LaneAir should consider these factors when interpreting the results.