

IDS 702 HW 5

Arko Bhattacharya

Guidelines: Follow these guidelines to finalize your assignment. Note that all code will remain hidden in the rendered document. Ensure you submit both this `.qmd` file and the corresponding rendered PDF. **Render your file frequently** to catch potential errors in your code ahead of time.

Update the Author section in the header with your name

Part 1: Data Science Team Report

Data Overview

The dataset comprises 3,478 records and 24 features, detailing passengers' flight experiences. Each record represents an individual passenger, including demographic, travel-specific, and satisfaction-related information. The dataset is free from missing values.

Key Attributes:

1. Demographics: Gender, Customer Type, Age
2. Flight Characteristics: Travel Type, Class, Flight Distance
3. Service Ratings (scale: 1–5): Including WiFi service, online booking, food, seat comfort, etc.
4. Delays: Departure and Arrival Delay in Minutes
5. Satisfaction: Categorized as “Satisfied,” “Neutral,” or “Dissatisfied”

Primary Insights:

- Service ratings show variability, with strengths in baggage handling and cleanliness
- Weak ratings in WiFi service and online booking processes
- High variability in flight delay times

Variable	Mean	SD
Age	39.67	15.11
Flight Distance	1186.76	984.99
Inflight Wifi Service	2.73	1.34
Ease of Online Booking	2.76	1.41
Gate Location	2.99	1.29
Food and Drink	3.22	1.33
Online Boarding	3.25	1.36
Seat Comfort	3.46	1.33
Departure Arrival Time Convenience	3.06	1.54
Inflight Entertainment	3.38	1.34
Onboard Service	3.41	1.28
Legroom Service	3.36	1.34
Baggage Handling	3.66	1.17
Check-in Service	3.32	1.26
Inflight Service	3.64	1.20
Cleanliness	3.29	1.32

Variable	Mean	SD
Departure Delay (min)	14.50	38.81
Arrival Delay (min)	15.20	39.37

Analysis Plan

Objectives:

1. Identify key factors influencing airline passenger satisfaction
2. Use two modeling approaches to understand satisfaction drivers

Modeling Approaches:

1. Ordinal Logistic Regression (OLR):
 - Handles ordinal response variables
 - Assumes consistent relationship between predictors and satisfaction
 - Uses logit link function
2. Multinomial Logistic Regression (MLR):
 - Treats satisfaction categories as independent
 - Provides more flexible modeling of satisfaction categories

Predictors Explored:

- Demographic Attributes
- Travel Features
- Service Evaluations
- Delays and Timeliness

These predictors were selected to capture a comprehensive view of the factors that may influence satisfaction, with both direct and indirect effects. By using both OLR and MLR models, we aim to identify the most significant drivers of satisfaction and provide actionable insights for improving the passenger experience.

Model Evaluation

Model performance will be assessed using accuracy, kappa, sensitivity, precision, and confusion matrix analysis. Additionally, we will evaluate the trade-off between interpretability (OLR) and flexibility (MLR) in identifying satisfaction drivers.

Model Results

Predictor	Odds Ratio	Lower CI	Upper CI	P-Value
Gender (Male)	0.99	0.86	1.15	0.938
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 Service	1.37	1.25	1.50	<0.001
Departure / Arrival Time Convenience	0.96	0.90	1.013	0.127
Customer Type (Loyal Customer)	5.24	4.17	6.60	<0.001
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	0.98	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
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 Findings from OLR Model:

- Customer Loyalty:
 - Loyal customers are 5.24 times more likely to report higher satisfaction

- Statistically significant ($P < 0.001$)
- Travel Type Impact:
 - Personal travel negatively affects satisfaction
 - Business travelers more likely to be satisfied
- Critical Service Quality Factors: Positive Predictors of Satisfaction:
 - Inflight WiFi (OR = 1.37) • Online Boarding (OR = 1.51) • Legroom Service (OR = 1.18) • Cleanliness (OR = 1.12)
- Minimal Demographic Influence:
 - Gender and age have little impact on satisfaction

Model Assessment

We compared predictions from the ordinal logistic regression model with those from the multinomial model. While the multinomial model predicts slightly higher probabilities for the dissatisfied category, especially for disloyal customers, the ranking of probabilities remains consistent. Differences in the neutral and satisfied categories are minor, suggesting no violation of the proportional odds assumption. Therefore, we find no evidence to reject the assumption.

Ordinal Logistic Model

- **Confusion Matrix**

Prediction / Actual	dissatisfied	neutral	satisfied
dissatisfied	502	511	76
neutral	300	277	114
satisfied	186	144	1368

- Key Metrics:
 - **Accuracy:** 61.73% – Above random chance, driven by strong performance in the “satisfied” category, but room for improvement.
 - **Kappa:** 0.4012 – Indicates moderate agreement between predictions and actual satisfaction levels.
 - **Balanced Accuracy:** 56.73% – Reflects challenges with underrepresented classes, like “neutral” customers.

Multinomial Model

- **Confusion Matrix**

Prediction / Actual	dissatisfied	neutral	satisfied
dissatisfied	464	423	116
neutral	341	368	61
satisfied	183	141	1381

- **Key Metrics:**

- **Accuracy:** 63.63% – The multinomial model outperforms the ordinal model, particularly in predicting the “satisfied” category.
- **Kappa:** 0.4309 – Shows moderate agreement, with better classification than the ordinal model.
- **Balanced Accuracy:** 61.85% – Enhanced sensitivity for neutral customers due to independent treatment of satisfaction levels.

Assessment Conclusion

The Multinomial Logistic Model outperforms the Ordinal Model in overall accuracy (63.63% vs. 61.73%) and balanced accuracy for neutral customers (61.85% vs. 56.73%), making it better at distinguishing satisfaction levels. However, the Ordinal Model offers greater interpretability with a simpler, more intuitive structure. For LaneAir’s strategy, where actionable insights and explainability are crucial, the ordinal model may still be the better choice despite its slightly lower predictive performance.

Part 2: Executive Summary

Introduction

This analysis examines survey data from 3,478 LaneAir passengers over the past six months, including demographic details, flight characteristics, service ratings, and satisfaction levels.

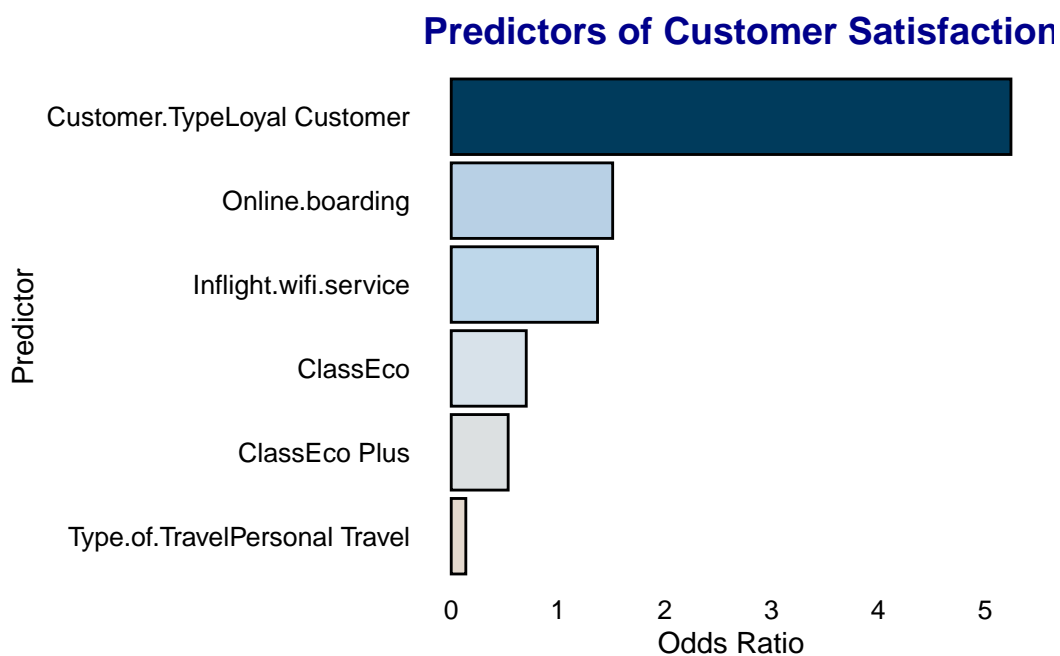
- **Satisfaction Levels:** 45% Satisfied, 27% Neutral, 28% Dissatisfied.
- **Service Improvement Areas:** Inflight WiFi (2.7/5), Online Booking (2.8/5).
- **Delays:** Avg. departure delay: 14.5 minutes, arrival delay: 15.2 minutes.

Methods

Two models were used:

- **Ordinal Logistic Regression (OLR):** Analyzes ordered satisfaction levels.
- **Multinomial Logistic Regression (MLR):** Identifies unique drivers for each satisfaction category.

Results



Key drivers of customer satisfaction were identified:

1. **Positive Predictors:**

- **Customer Loyalty:** Loyal customers are 5.24 times more likely to be satisfied than disloyal customers.
- **Online Boarding:** Positive ratings for online boarding increase satisfaction odds by 1.5 times.

2. **Negative Predictors:**

- **Personal Travel:** Passengers traveling for personal reasons are less likely to report satisfaction compared to business travelers.
- **Eco Plus and Eco cabin class:** The odds of an Eco Plus passenger being satisfied are 47% lower compared to passengers who are not in this section. Similarly, the odds of an Eco passenger being satisfied are 30% lower compared to passengers who are not in this section.

Conclusion

To enhance customer satisfaction, LaneAir should prioritize the following:

1. **Enhance Customer Loyalty Programs:** Loyal customers are significantly more likely to be satisfied. Expanding loyalty programs can boost satisfaction and retention.
2. **Optimize Online Boarding and Inflight Services:** Improving online boarding and inflight WiFi services can streamline customer experiences and increase satisfaction at key touchpoints.
3. **Improve Legroom and Seat Comfort:** While seating changes may incur costs, small adjustments to comfort and legroom can have a substantial impact on satisfaction.
4. **Improve Operational Timeliness:** Timely arrivals are essential for maintaining customer satisfaction, particularly for time-sensitive passengers. Minimizing delays should be a priority.
5. **Address Eco and Eco Plus Cabin Classes:** Passengers in the Eco and Eco Plus cabins report lower satisfaction. LaneAir should consider enhancing the service and amenities in these lower-cost cabins to improve satisfaction levels.

Additional Considerations:

LaneAir should prioritize cost-effective strategies such as enhancing online boarding and in-flight services, which offer significant improvements without major investments. In addition, improving staff training, particularly in the Eco and Eco Plus cabins, can elevate the customer experience without the need for costly infrastructure changes. It's also important to note that the survey may be biased toward more frequent or business travelers, potentially overlooking the needs of other customer segments. Further research with a more diverse sample could offer more comprehensive insights and help refine satisfaction improvement strategies.