IDS 702 HW 5

Eric Ortega Rodriguez

Part 1: Data Science Team Report

Data Overview

The dataset comprises 3,478 observations and 24 variables. It was collected through a customer satisfaction survey for an airline, with the primary goal of identifying key factors that influence customer satisfaction. Each row represents a unique passenger, and the columns include a variety of characteristics such as demographics, travel-related details, service satisfaction, and delay-related features. The dataset is complete, with no missing values, simplifying the data preparation process.

The satisfaction levels among passengers are distributed as follows: - **Satisfied**: 44.8% (1,558 passengers) - **Neutral**: 26.8% (932 passengers) - **Dissatisfied**: 28.4% (988 passengers)

Below is a detailed breakdown of the dataset's attributes:

1. **Demographics**:
    - **Gender**: Passenger's gender (e.g., Male, Female).
    - **Customer.Type**: Indicates if the passenger is a Loyal or Disloyal Customer.
    - **Age**: Passenger's age.
2. **Flight Characteristics**:
    - **Type.of.Travel**: Purpose of travel (Business or Personal).
    - **Class**: Flight class (e.g., Business, Economy, Economy Plus).
    - **Flight.Distance**: Total flight distance in miles.
3. **Service Ratings** (scale: 1–5, where 5 indicates the highest satisfaction):
    - Includes `Inflight.wifi.service`, `Ease.of.Online.booking`, `Food.and.drink`, `Seat.comfort`, `Inflight.entertainment`, `On.board.service`, `Leg.room.service`, `Baggage.handling`, `Checkin.service`, `Inflight.service`, and `Cleanliness`.
4. **Delays**:
    - **Departure.Delay.in.Minutes**: Duration of delay at departure.
    - **Arrival.Delay.in.Minutes**: Duration of delay at arrival.

This dataset provides a robust foundation for analyzing factors that drive customer satisfaction and identifying opportunities for improvement in the airline's operations.

The table below summarizes the descriptive statistics for the numeric variables in the dataset, including the mean and standard deviation.

Table 1: Descriptive Statistics of Numeric Variables

Variable	Mean	Standard Deviation
Passenger Age	39.673950	15.110163
Flight Distance (miles)	1186.756469	984.994848
Inflight Wifi Service Rating	2.731742	1.337789
Ease of Online Booking Rating	2.761645	1.410079
Gate Location Rating	2.988212	1.289603
Food and Drink Rating	3.219091	1.333519
Online Boarding Rating	3.251869	1.357068
Seat Comfort Rating	3.461185	1.328030
Inflight Entertainment Rating	3.380104	1.341523
Onboard Service Rating	3.409431	1.282946
Legroom Service Rating	3.355089	1.335180
Baggage Handling Rating	3.663887	1.167837
Check-in Service Rating	3.323174	1.263723
Inflight Service Rating	3.637148	1.196303
Cleanliness Rating	3.294422	1.323067
Departure Delay (minutes)	14.495687	38.806258
Arrival Delay (minutes)	15.196665	39.371572

Analysis Plan

Our analysis aims to identify the key factors influencing customer satisfaction and provide actionable insights for LaneAir. To achieve this, two models will be implemented to address the ordinal nature of the target variable (`Satisfaction`) and ensure robustness in the findings.

Type of Models

The following models will be used for the analysis: 1. **Ordinal Logistic Regression**: - This is the primary model as it will look at ordinal target variables (`Satisfaction`) with three levels: `Dissatisfied`, `Neutral`, and `Satisfied`. This model assumes that the relationship between each predictor and the outcome is consistent across all thresholds of satisfaction. 2. **Multinomial Logistic Regression**: - This model is used to compare with the ordinal logistic regression and validate its assumptions. Unlike the ordinal model, it treats `Satisfaction` as a nominal variable, allowing for more flexibility in capturing relationships between predictors and satisfaction levels.

Link Function

1. **Ordinal Logistic Regression**:
    - The **link function** will be used. It models the cumulative probabilities of satisfaction levels and assumes proportional odds among variables.
2. **Multinomial Logistic Regression**:
    - The **link function** will also be used for this model. However, it does not assume proportional odds and instead models each satisfaction level independently. Essentially, it will be its own category.

Predictors

The predictors used in this analysis are grouped into four general categories: 1. **Demographics**: - Variables: `Age`, `Gender`, `Customer.Type`.

2. **Travel Characteristics**:
    - Variables: `Type.of.Travel`, `Class`, `Flight.Distance`.
    - Capture trip and the service class that are linked to passenger expectations and satisfaction.
3. **Service Ratings**:
    - Variables: Ratings for `Inflight.wifi.service`, `Ease.of.Online.booking`, `Seat.comfort`, `Food.and.drink`, `Online.boarding`, `Inflight.entertainment`, `Cleanliness`, `On.board.service`, `Leg.room.service`, `Baggage.handling`, `Checkin.service`, and `Inflight.service`.
    - Capture passenger feedback on various aspects of inflight and airport services.
4. **Operational Factors**:
    - Variables: `Departure.Delay.in.Minutes`, `Arrival.Delay.in.Minutes`, `Departure.Arrival.time.convenient`.
    - Capture the performance of the airline which significantly impact satisfaction.

Model Evaluation Plan

Model performance will compare the Ordinal Logistic Regression (OLR) and Multinomial Logistic Regression (MLR) models in determining customer satisfaction. Evaluation metrics will include overall accuracy, assessed against the No Information Rate, and the kappa statistic. A confusion matrix will be used to identify issues and differences among the models. The OLR model will be examined for its interpretability, while the MLR model will be evaluated for its

flexibility in handling non-ordinal relationships. Finally, the Akaike Information Criterion (AIC) will compare model fit.

Model Results

Table 2: Ordinal Logistic Regression Results

Predictor	Odds Ratio	Lower CI	Upper CI	P-Value
Gender (Male)	0.994	0.859	1.151	0.938
Age	0.995	0.990	1.000	0.034
Type of Travel (Personal)	0.137	0.110	0.170	<0.001
Class (Economy)	0.703	0.575	0.859	<0.001
Class (Economy Plus)	0.533	0.396	0.717	<0.001
Flight Distance	1.000	0.999	1.001	0.742
Inflight Wifi Service	1.370	1.251	1.501	<0.001
Departure / Arrival Time Convenience	0.956	0.903	1.013	0.127
Customer Type (Loyal Customer)	5.242	4.169	6.602	<0.001
Ease of Online Booking	0.747	0.684	0.816	<0.001
Gate Location	0.992	0.927	1.061	0.802
Food and Drink	0.978	0.898	1.062	0.589
Online Boarding	1.512	1.401	1.633	<0.001
Seat Comfort	1.057	0.973	1.143	0.191
Inflight Entertainment	1.056	0.950	1.174	0.3
Onboard Service	1.165	1.085	1.246	<0.001
Legroom Service	1.179	1.111	1.247	<0.001
Baggage Handling	1.079	0.992	1.173	0.069
Check-in Service	1.209	1.136	1.283	<0.001
Inflight Service	1.133	1.042	1.223	0.003
Cleanliness	1.118	1.015	1.230	0.022
Departure Delay in Minutes	1.005	0.998	1.012	0.199
Arrival Delay in Minutes	0.993	0.986	1.000	0.042

Based on the chart above we can see that customer loyalty, travel type, class, and certain service-related factors. Such facors include wifi and online booking, are the most compelling drivers of customer satisfaction. I would advise the airline to targeting these areas as they are most likely to improve overal customer satisfaction.

Model Comparison

We will now compare the Ordinal Logistic Regression (OLR) model to the Multinomial Logistic Regression (MLR) model to determine which approach is more effective in identifying the factors contributing to customer satisfaction.

The following tables compare the **Ordinal Logistic Regression (OLR)** and **Multinomial Logistic Regression (MLR)** models:

Confusion Matrix for Ordinal Logistic Regression (OLR)

Prediction	Dissatisfied	Neutral	Satisfied
Dissatisfied	502	511	76
Neutral	300	277	114
Satisfied	186	144	1368

Confusion Matrix for Multinomial Logistic Regression (MLR)

Prediction	Dissatisfied	Neutral	Satisfied
Dissatisfied	464	423	116
Neutral	341	368	61
Satisfied	183	141	1381

Model Comparison Summary

Model	AIC	Accuracy	Kappa
OLR	5516.347	0.6173	0.4012
MLR	5006.669	0.6363	0.4309

Based on the results above , the Multinomial Logistic Regression model shows a better fit with a lower AIC value (5006.67) compared to the Ordinal Logistic Regression model (5516.35). Additionally, the Multinomial Logistic Regression model has a higher accuracy (63.63%) and kappa (0.43), indicating better performance in predicting customer satisfaction.

On the other hand, the Ordinal Logistic Regression model would also be a good option due to its simpler structure. In conclusion, the MLR model is effective in capturing nuances within the data.

Part 2: Executive Summary

Introduction

The LaneAir customer satisfaction dataset comprises 3,478 survey responses with 24 variables.

The varaibles range from demographics (gender, age, customer type), travel characteristics (such as travel purpose, flight class, flight distance), service ratings (such as inflight WiFi, food and drink, seat comfort), and to operational factors (such as delays).

The primary goal of this analysis is to identify key factors influencing customer satisfaction to recommend actionable improvements.

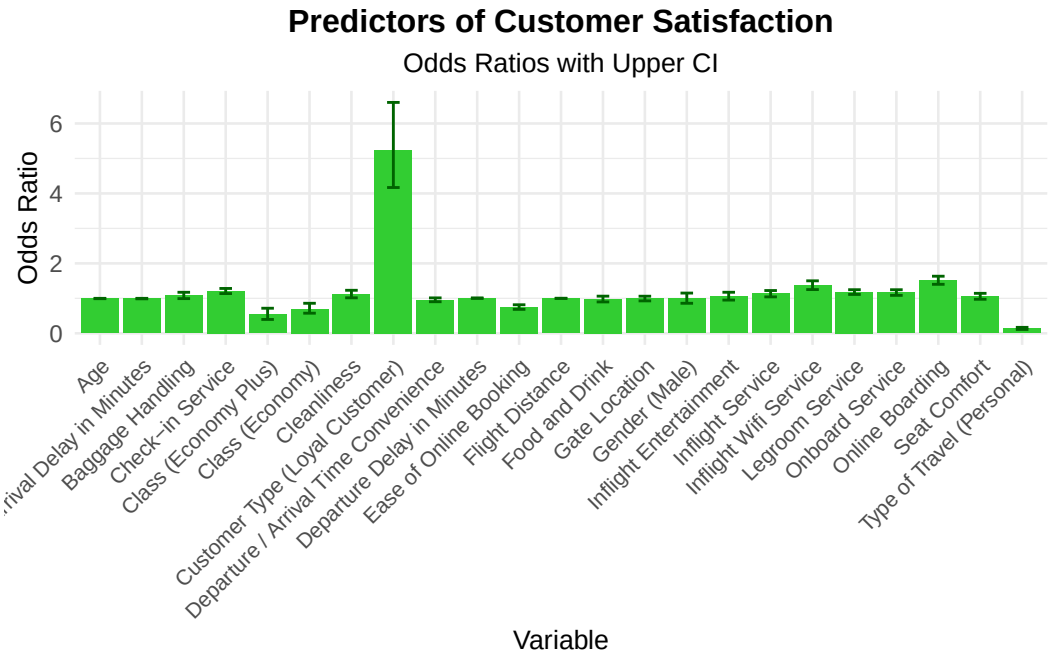
Methods

To analyze customer satisfaction, two models were implemented: 1. **Ordinal Logistic Regression (OLR):** This model assumes an ordinal relationship between predictors and satisfaction levels (dissatisfied, neutral, satisfied). 2. **Multinomial Logistic Regression (MLR):** A model that treats satisfaction levels as independent categories providing detailed insights.

Results

- **Customer Loyalty and Travel Type:** Loyal customers and business travelers show significantly higher satisfaction odds.
- **Service Quality:** High satisfaction levels correlate strongly with WiFi service, online boarding, and inflight service.
- **Operational Factors:** Delays impact satisfaction minimally, likely due to customer expectations.

Model comparison revealed the **MLR model** performed better, with higher accuracy (63.63%) and a lower AIC (5006.67) than the OLR model.



Conclusion

To enhance customer satisfaction effectively and cost-efficiently: 1. Invest in **Technology that will improve wifi and entertainment service** for immediate impact. This is an option that is more plausible for Laneair.

2. Improve **customer loyalty programs and marketing**, as through our results, we found that loyal customers are are up to five times more likely to be satisfied with their flight.
3. Ulitize **online booking, boarding processes, technologies that will improve booking speeds** which will improve convienience for the costomer and overall satisfaction.

Limitations: The survey responses may have non-response bias, with disloyal customers less likely to participate. Also the satisfaction drivers vary across customer demographics.

If Laneair focuses on service improvements and leveraging customer insights, it can improve overall customer satisfaction. This is also more align with cost management since they are more plausible.