

Predicting and Analyzing British Airway airplane tickets purchase

Data Overview:

- The dataset contains 50,000 entries with 14 columns related to customer flight bookings

- No missing values found in any columns

- Data types include integers (8), floats (1), and objects (5)

- Columns include booking details like passenger count, sales channel, trip type, flight details, and whether booking was completed

Key Statistics:

- Average booking completion rate: 14.96%

- Average purchase lead time: 84.94 days

- Average flight duration: 7.28 hours

- Most bookings are made via Internet (sales\_channel)

- Most common trip type is RoundTrip (trip\_type)

- Most common booking origin is Australia (booking\_origin)

Visualizations Insights:

1. Violin Plots show distributions of numerical variables:

- Most flights have 1-2 passengers (num\_passengers)

- Purchase lead time is right-skewed with some very early bookings (up to 867 days)

- Flight hours show peaks around morning (7am) and evening (3pm, 5pm)

- Most flights are on weekends (flight\_day mapped to Sat=6, Sun=7)

2. Categorical Data:

- Internet is the dominant sales channel

- Round trips are most common

- Top routes include AKLDEL and PERPNH

Model Performance Considerations

The notebook shows exploratory analysis but doesn't yet include predictive modeling. Based on the data:

1. Target Variable: booking\_complete (binary classification)

- Imbalanced dataset (only 14.96% positive class)

- May require techniques like SMOTE oversampling/undersampling

2. Key Features that likely impact booking completion:

- Purchase lead time (longer leads may cancel more)

- Flight hour/day (timing may affect completion)

- Route and booking origin (geographical patterns)

- Extra services requested (baggage, meals, seats)

3. Applied Machine Learning Models:

- Logistic Regression (baseline)

- Random Forest (handles imbalanced data well)

- XGBoost (good for tabular data)

- Need to evaluate precision/recall given class imbalance

4. Feature Engineering Opportunities:

- Create time-based feature Late\_Booking from purchase\_lead

- Group less common routes/origins

- Combine service preferences into single feature