Drivers of Booking completions

Name : Imanuel Annoh

Problem Statement

- Goal of the project is to predict whether an airline passenger would complete their flight booking, enabling targeted marketing and better resource allocation

Data Overview

- Internal source of data from British Airways past consumer behavior on whther an airline passenger completed thier booking or not
- Key columns (target: booking_complete, input features like route, booking_origin, preferences)
- They were around 50,000 records available in the datasets and 14 feature columns

Data preprocessing

- No missing values present
- Categorical columns were encoded to numerical columns using the pandas get dummies method
- Transformed the high cardinal features route, and booking_origin using feature encoding to reduce the the dimensionality space

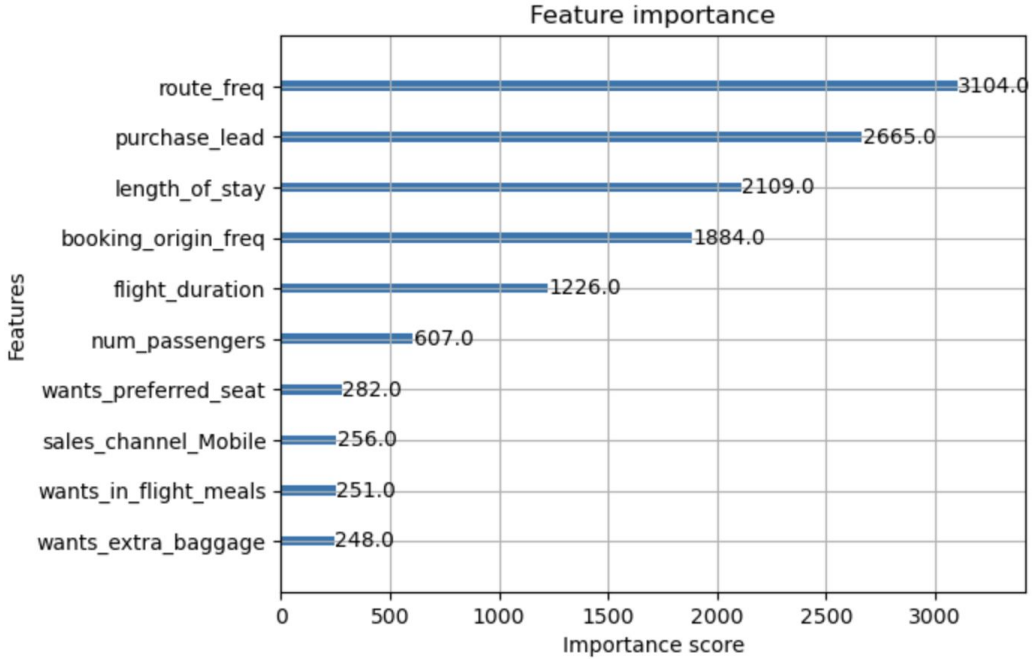
Exploratory Data Analysis

- Chi Square test and Point biserial correlation performed to measures levels of association with the target variable
- While both correlational analysis showed statistical relationships between the features and target variable, the strength of these correlations were broadly weak

Model Training and Evaluation

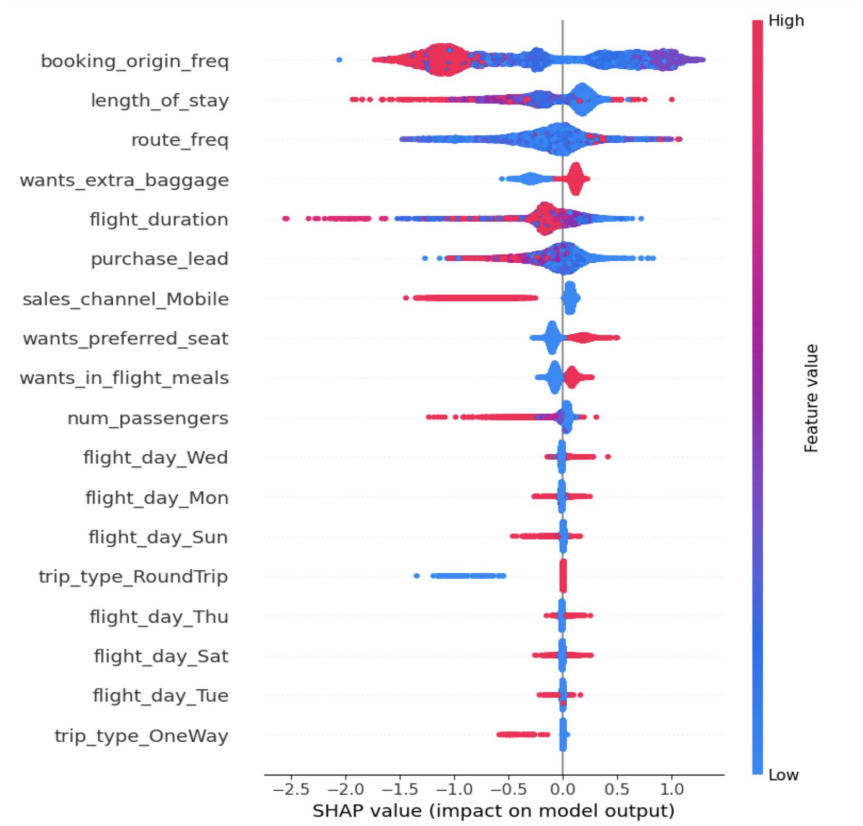
- Three binary classifiers (logistic regression, Random Forest Classifier , XGBoost Classifier) were trained on the dataset
- Initial model evaluation from the three classifiers showed XGboost as the most promising model with the highest recorded F1-score (0.44)
- Hyperparameter tuning was performed on the XGBoost classifier and metrics then evaluated
- Tuned XGBoost classifier showed modest gains in metric evaluation

Model Interpratability



- The most important predictors of booking completion were route_freq, purchase_lead, and length_of_stay, suggesting that route popularity and traveler planning behavior are key indicators. Passenger preferences such as wanting meals or extra baggage also contributed modestly

SHAP analysis



- Top predictors include booking_origin_freq, length of stay
- From the SHAP plot, the higher a passenger wanted extra baggage the more likely it was they completed the completed
- Also, low feature values of booking origin freq for a passenger meant they were also likely to complete the booking process