Analysis of Customer Booking Predictions

Summary of Data Exploration, Model Training, and Evaluation

Dataset Overview

- 14 columns
- 50k entries
- No missing values

Data Preprocessing

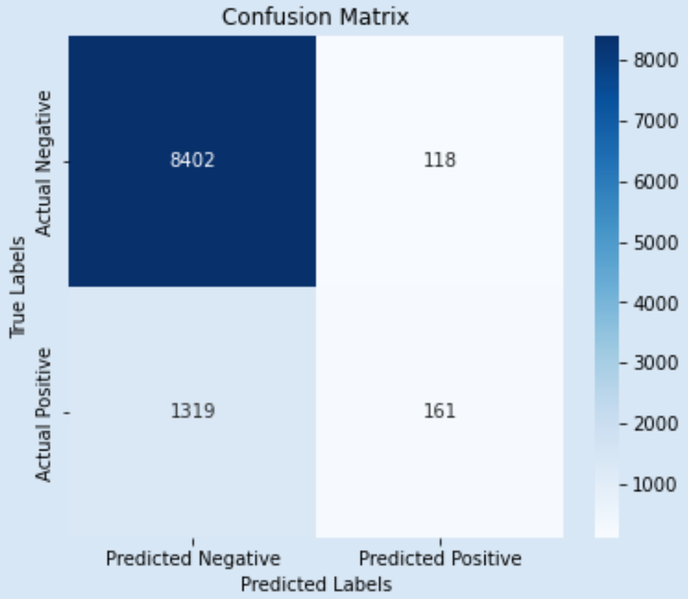
- Applied label encoding for columns: 'sales_channel', 'trip_type', 'flight_day', 'route', 'booking_origin'

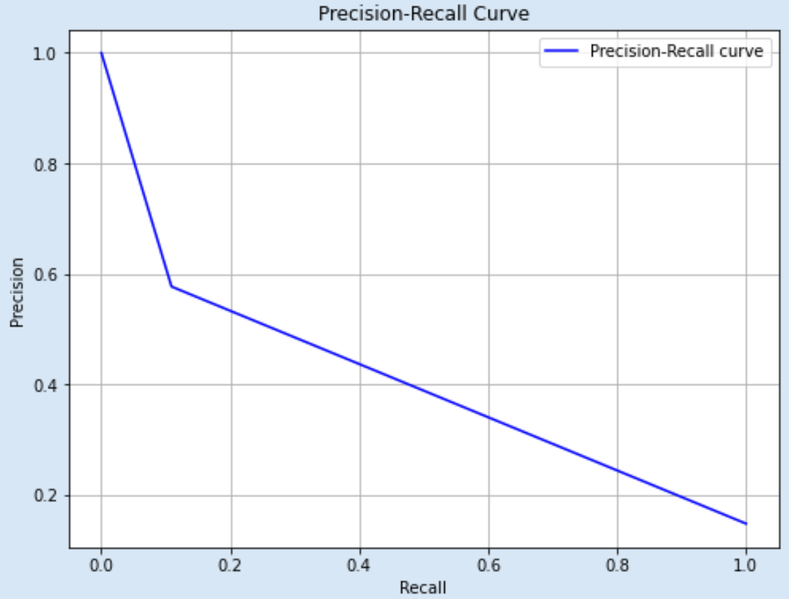
Machine learning model

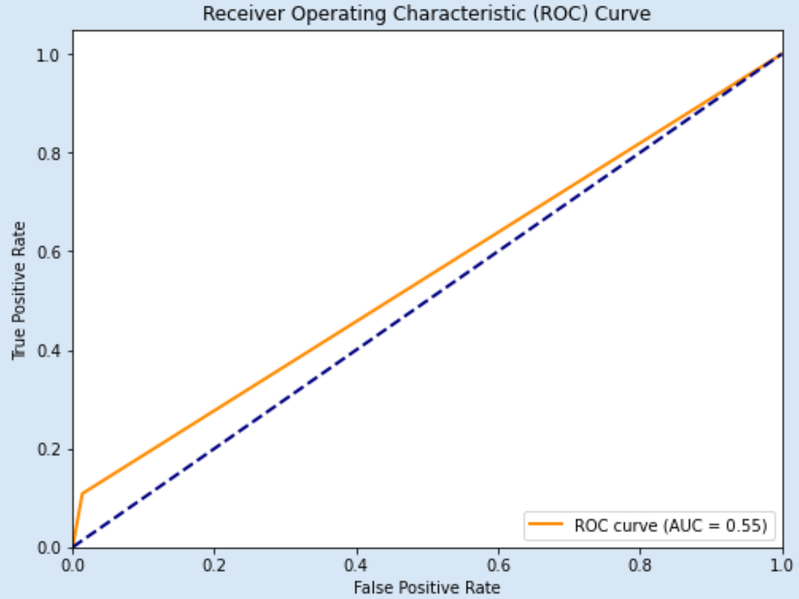
- Model used: Random Forest Classifier
- Train-Test Split Ratio: 80:20

Model Evaluation

- Accuracy: 0.8563
- Recall: 0.1088
- Precision: 0.5771
- F1-Score: 0.1831



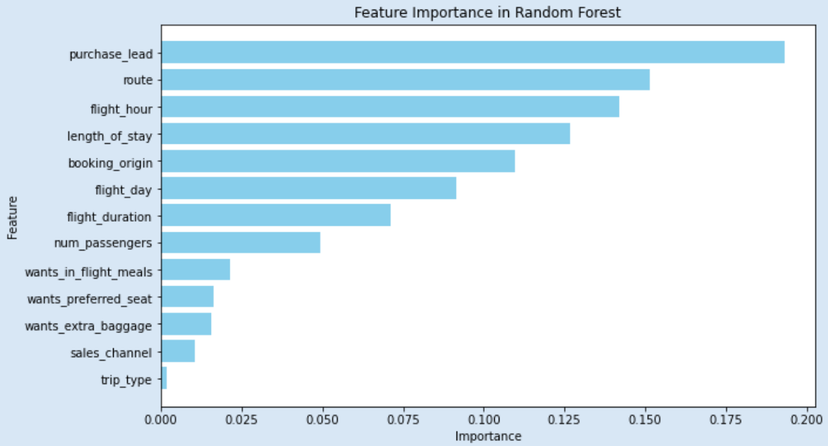




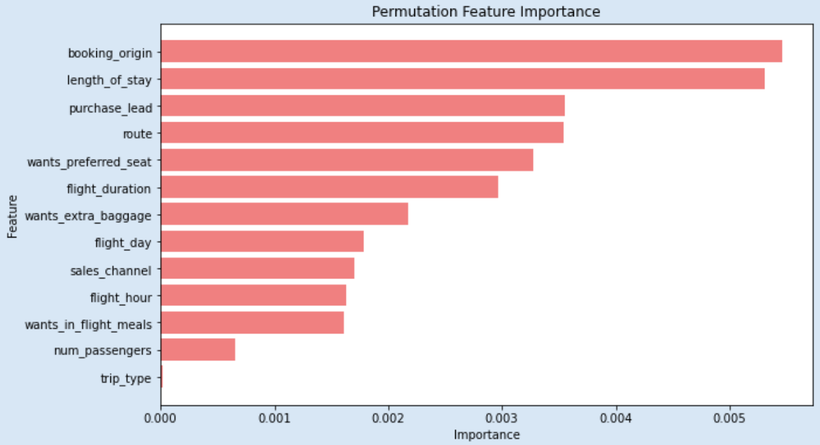
Analysis of Customer Booking Predictions

Summary of Data Exploration, Model Training, and Evaluation

Feature Importance



Permutation Importance



Insights

A) **purchase_lead** is the most influential feature in the booking model prediction

reasons:

- booking behaviour patterns
- price sensitivity
- flight demand

B) **booking_origin** is the most significant feature that impacts the model's predictive accuracy

reasons:

- geographical factors
- economic factors
- targetted offers and promotions

C) The model is cautious in predicting positives, possibly due to an **imbalanced dataset** or the high cost of false positives.

reasons:

- high precision and low recall
- AUC 0.55