BRITISH ARIWAYS

CUSTOMER BOOKING MODEL

a presentation from data scientsi at British Airways

www.britishairways.com

BRITISH AIRWAYS PRESENTATION

Table of Content

ABOUT COMPANY

DATA UNDERSTANDING

EXPLORATORY DATA

DATA PREPROCESSING

MODELLING

this table of content, contains some information about the presentation into a few pages on how we create a plan or strategy model for a predictive customer booking using machine learning, with other information too like the feature importance and etc.

www.britishairways.com

About Company

Established in 1974, British Airways stands as a globally recognized flag carrier renowned for its extensive route network, exemplary service, and commitment to innovation. Rooted in the legacy of pioneering airlines like Imperial Airways and BOAC, British Airways operates from London Heathrow Airport, connecting over 200 destinations worldwide across Europe, North America, Asia, and Africa. Embodying sophistication and comfort, the airline prioritizes customer satisfaction through luxurious first-class suites and efficient economy cabins. Committed to sustainability, it invests in fleet modernization and environmental initiatives, while actively engaging in corporate social responsibility endeavors.

Data Understanding

50.000

Passenger on our dataset

14

columns storing various booking and preference information.

Comprehensive data coverage for diverse analysis.

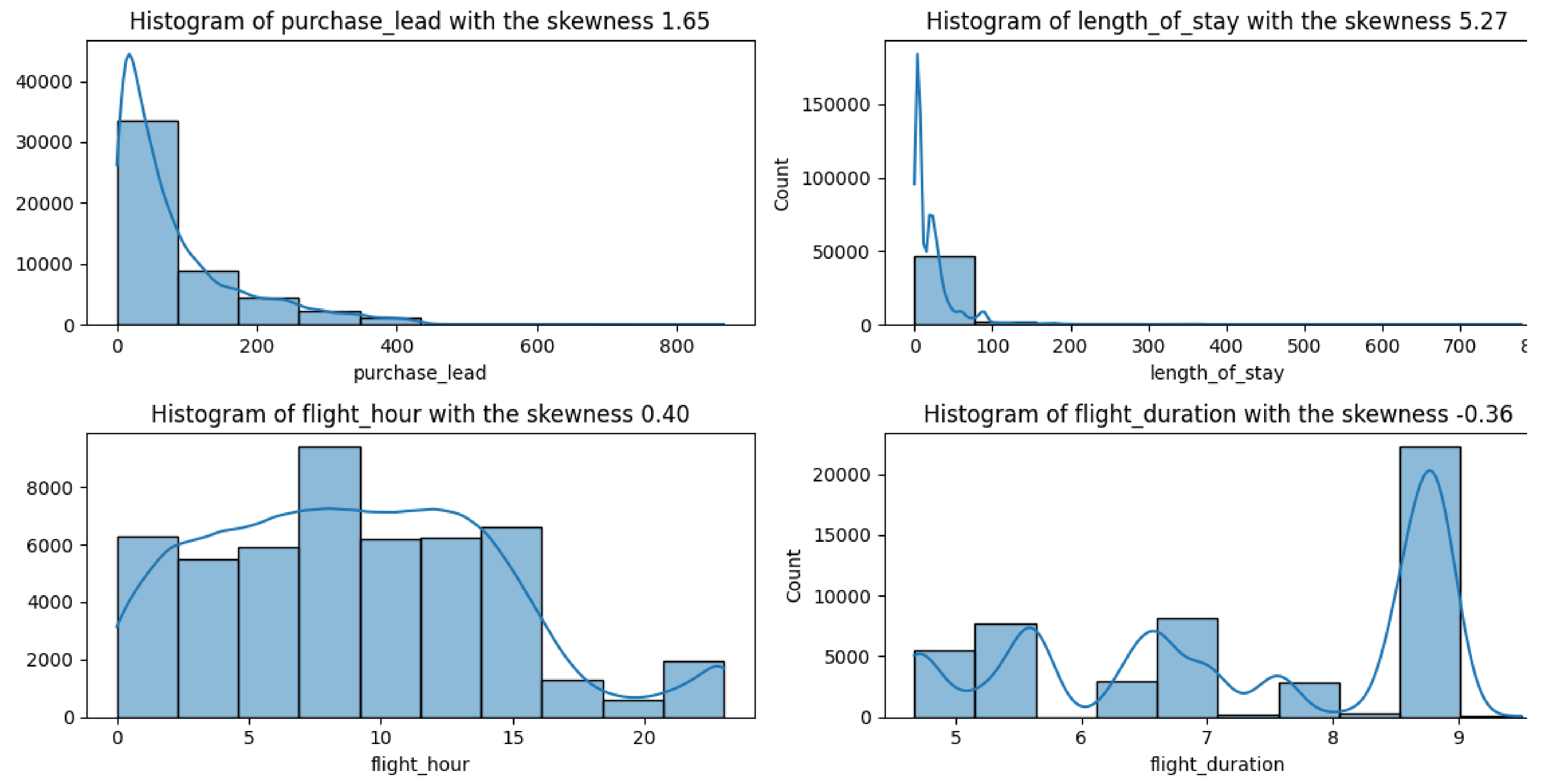
No missing values observed across all entries.

Exploratory Data Analysis

Purchase lead time and length of stay display significant skewness in their distributions.

Flight duration and flight hour exhibit moderate skewness, indicating a relatively balanced distribution around their means.

Handling or transforming highly skewed features like purchase lead time and length of stay may be necessary. Flight duration and flight hour, though moderately skewed, present less extreme variations in the data.



Data Preprocessing

REMOVE DUPLICATE DATA

To ensure data integrity, several duplicate rows were identified and removed, streamlining the dataset for analysis by eliminating redundant entries.

FEATURE ENGINEERING

A new feature, **'purchase_lead_length_of_stay_interaction,'** and **'flight_hour'** to make the data more diverse.

SCALING NUMERICAL

Utilizing the MinMaxScaler, numerical features were scaled, normalizing their values within a specified range.

ENCODING DATA

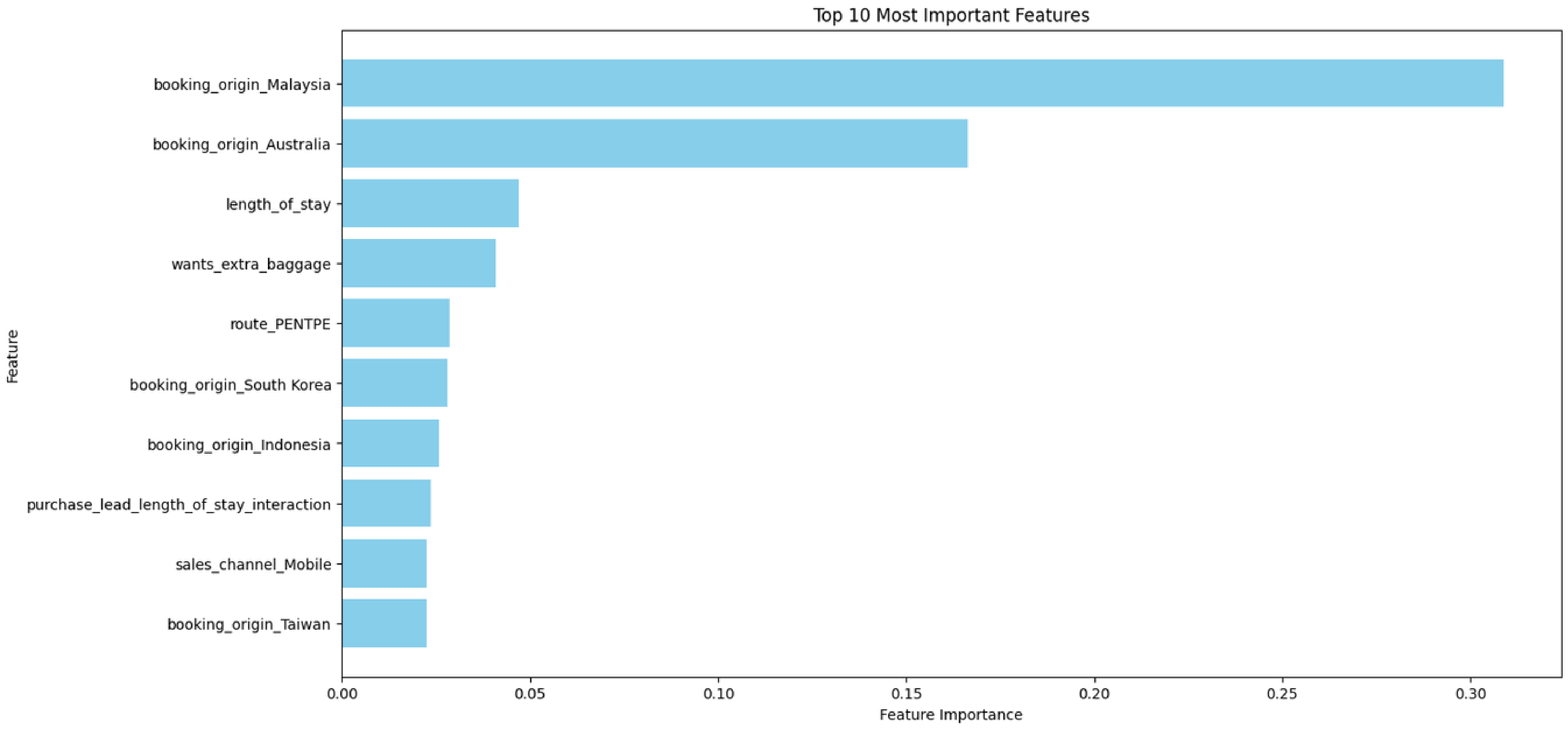
Employing one-hot encoding, all categorical data underwent transformation,

Model Comparison

Gradient Boosting was selected for its balanced performance across metrics, exhibiting better precision and F1-score for class 1 predictions. Despite a slightly lower recall, it demonstrated nuanced capabilities in handling both classes effectively.

Model	Accuracy	Precision (Class 1)	Recall (Class 1)	F1-Score (Class 1)
Random Forest	0.85	0.51	0.10	0.17
Logistic Regression	0.85	0.42	0.07	0.11
Gradient Boosting	**0.85**	**0.58**	**0.03**	**0.05**

feature importance



The top three influential features in our analysis are 'booking_origin' for Malaysia, 'booking_origin' for Australia, and 'length_of_stay.' These features played a significant role in the predictive capacity of our model, contributing notably to the outcomes and insights drawn from our analysis.

✨Thank you for Participating✨