BRITISH AIRWAYS

BRITISH AIRWAYS

Data Analysis

TABLE OF **CONTENTS**

ABOUT US

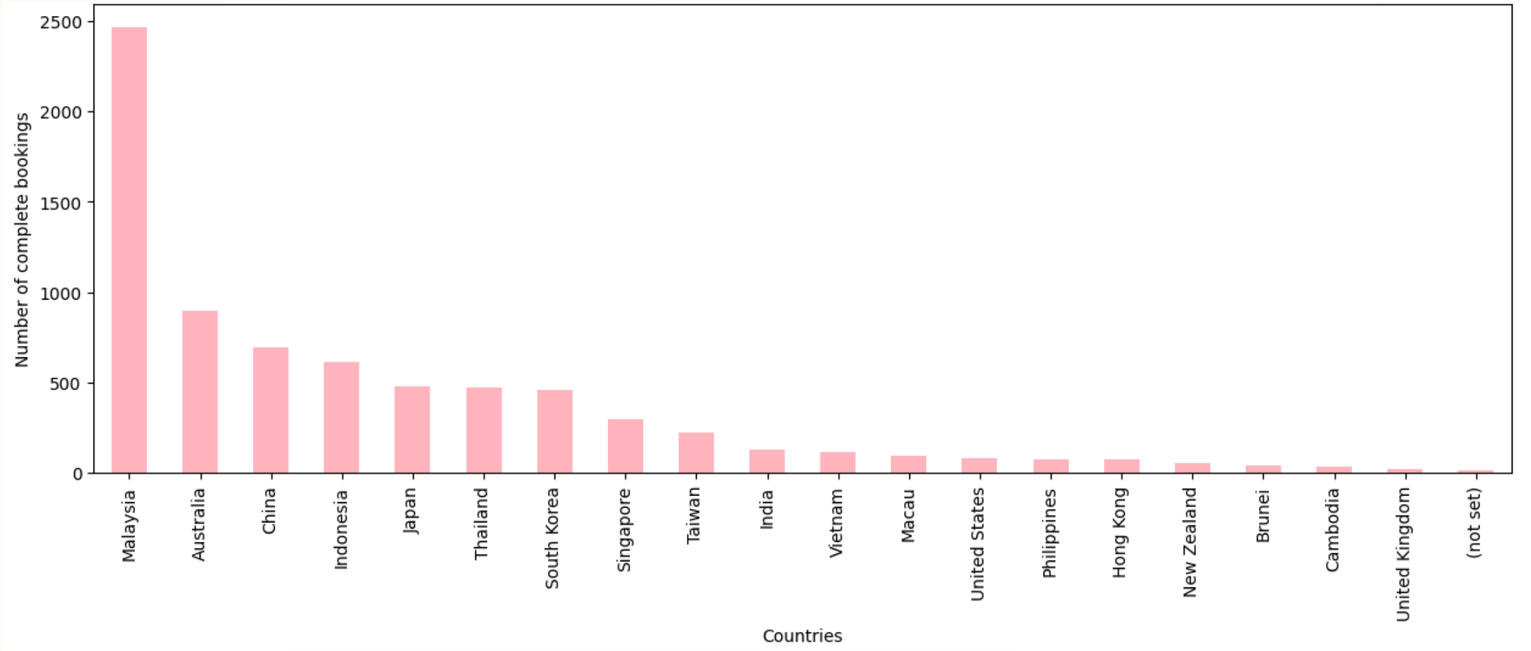
British Airways is the UK's flagship carrier, serving over 180 destinations in 80+ countries. The airline offers a range of travel classes and is committed to exceptional customer service. British Airways is a founding member of the oneworld alliance and is focused on sustainability, aiming for net-zero carbon emissions by 2050.

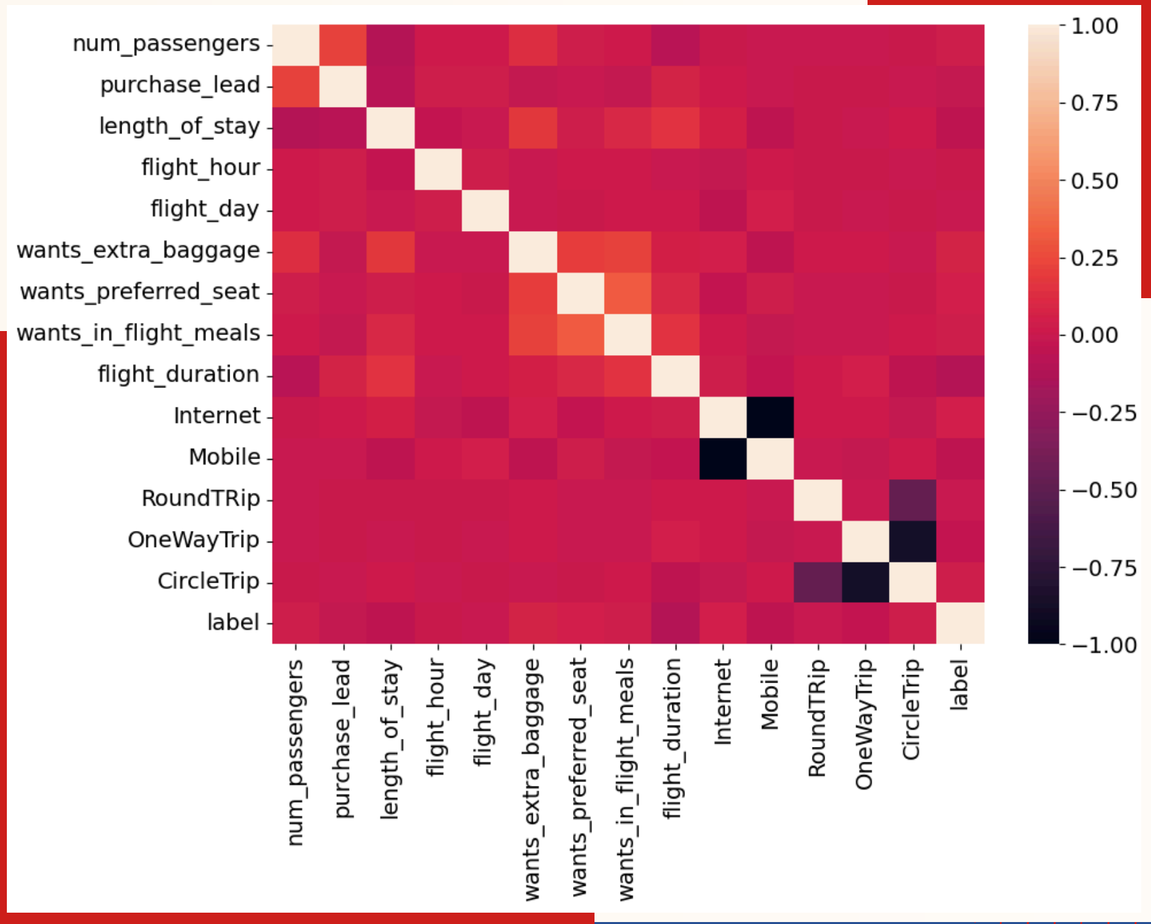
PROPOSED **OBJECTIVES**

- **Explore and understand the dataset:** Perform an initial exploratory data analysis to understand the structure, key features, and statistical properties of the dataset.
- **Data preparation and feature engineering:** Clean and preprocess the dataset to handle missing values, normalize data if needed, and create new features that could improve the model's predictive accuracy.
- **Model training and prediction:** Develop a machine learning model (e.g., RandomForest) to predict whether a customer will make a booking, focusing on optimizing accuracy and interpretability.
- **Evaluate and interpret model performance:** Assess the model's performance using metrics such as accuracy, precision, recall, and F1-score, and create visualizations to interpret feature importance.

COMPLETE **BOOKINGS**

- This bar chart displays the 19 countries that had British Airways complete bookings.
- Malaysia being the top most with complete bookings.
- Out of 50000 booking entries only 14.96 % bookings were successful or complete.



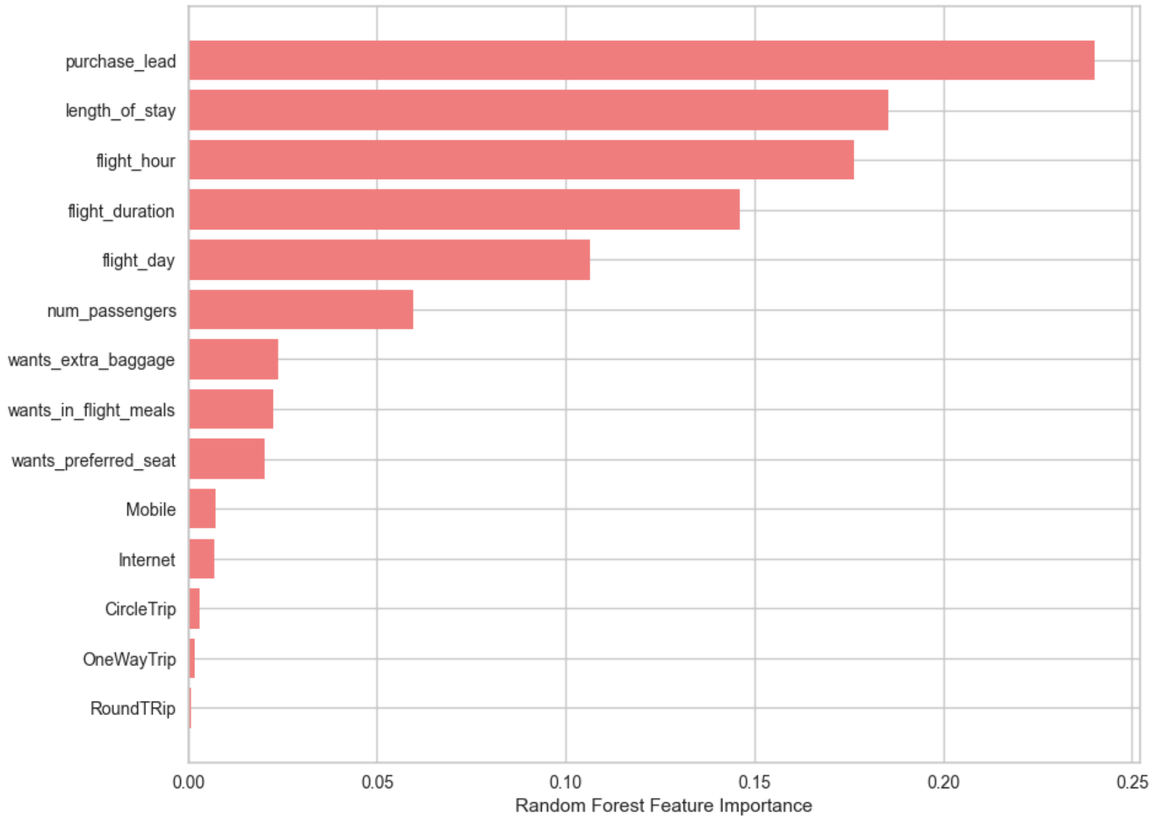


HEAT MAP

This heatmap illustrates the correlation matrix of various features from the dataset, where the color intensity represents the strength and direction of the correlation. Features with a correlation close to 1 have a strong positive relationship, indicated by lighter shades, while those with a correlation close to -1 have a strong negative relationship, shown in darker shades.

IMPORTANT **SCORES**

This bar chart represents the feature importance scores from the RandomForest model used for predicting customer bookings. The *purchase_lead* feature is the most influential, contributing the highest to the model's predictive power. Other important features include *length_of_stay*, *flight_hour*, *flight_duration*, and *flight_day*, which also have significant importance. Features like Mobile, Internet, and OneWayTrip have much lower importance scores, indicating they contribute minimally to the model's predictions.



PREDICTIVE MODELING

Recall rate= 59%

The likelihood of the model accurately identifying successful bookings.

Accuracy= 61%

The model's overall effectiveness in predicting both successful and incomplete bookings.

Precision= 59%

The proportion of correctly predicted completed bookings among all actual completed bookings.

CONCLUSION

The project successfully developed a predictive model using the RandomForest algorithm to forecast customer bookings. After performing feature engineering and model evaluation, the model achieved an accuracy of 0.61, precision of 0.59, recall of 0.59, and an F1-score of 0.62 on the training data. Key factors influencing the predictions included purchase lead, length of stay, and flight hour. These insights, along with feature importance visualizations, provide actionable information for improving customer booking strategies and resource allocation.