

# Challenge Description

Project: Hotel Booking Cancellation Prediction

## Task

You are provided with a dataset of hotel bookings. The goal is to predict whether a booking will be canceled based on various features such as lead time, number of guests, meal preferences, and others.

Your task is to build a classification model to predict the target variable **is\_canceled**, which indicates whether a booking was canceled (1) or not (0).

## File Descriptions

1. **train.csv** - The training set containing hotel booking data with labels (is\_canceled).
2. **test.csv** - The test set for which predictions are required. This file does not include the is\_canceled column.
3. **sample\_submission.csv** - A sample submission file showing the required format of predictions.

## Submission File

For each booking in the test set, you must predict the probability of the booking being canceled. The submission file should contain a header and be in the following format:

```
id,is_canceled
1,0.65
2,0.25
3,0.90
```

Where:

- id corresponds to the booking ID.
- is\_canceled is the predicted probability of cancellation.

## Evaluation Metric

Submissions will be evaluated based on the **F1 score**. Ensure your model achieves a balance between precision and recall to maximize the F1 score.