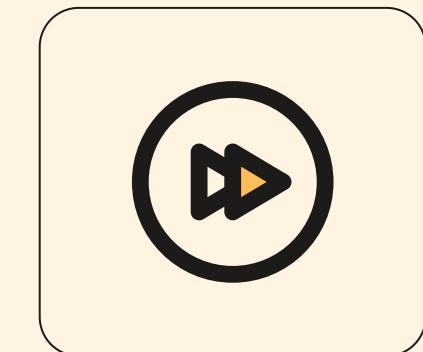


# DATA ANALYTICS

PREDICT HOTEL BOOKING  
CANCELLATION RATE AND  
DEVELOP DYNAMIC CUSTOMER  
RETENTION STRATEGIES

Ying Chu  
Ganapathi Teja Chirravuri  
Ahmed Al-Zubaidi



# MOTIVATION



## Revenue Loss

- Industry average cancellation rates range 37% (according to kaggle data)
- lead to empty rooms and misallocation of human resources



Solution

Shifting from reactive handling to proactive prevention

# GOAL



- Predict the probability of booking cancellation
- Identify the Top 5 factors influencing cancellation

# DATASET

Total Records : 119,390

Data range : 01/07/2015 - 21/08/2017

Columns : 31

Target Variable : is\_canceled

---

Cancellation Rate : 37%

>> Imbalanced Data

prioritize the f1-Score as the primary metric and use  
model-based class weighting to manage data imbalance

**Hotel & Booking  
Information**

**Room & Booking  
Changes**

**Stay Details**

**Agent & Company**

**Meal & Market  
Information**

**Customer & Pricing**

**Guest Behavior**

**Reservation Status**

# METHODOLOGY & MODEL STRATEGY

## Data Preprocessing

- Drop columns with a high proportion of missing values.
- Create time-related features .
- Apply target encoding to categorical variables.
- Check for multicollinearity.
- Remove highly correlated variables.
- Categorical encoding.

## Imbalance Handling and Weight Adjustment

# METHODOLOGY & MODEL STRATEGY

## Model Selection

- **baseline : Logistic Regression**
- **main model : Random Forest / XGBoost / LightGBM / Voting Classifier**

## Focus on F1-Score & AUC-ROC

- **It better reflects the model's capability to predict the minority class of cancellation.**

# PROJECT TIMELINE

# **THANK YOU**