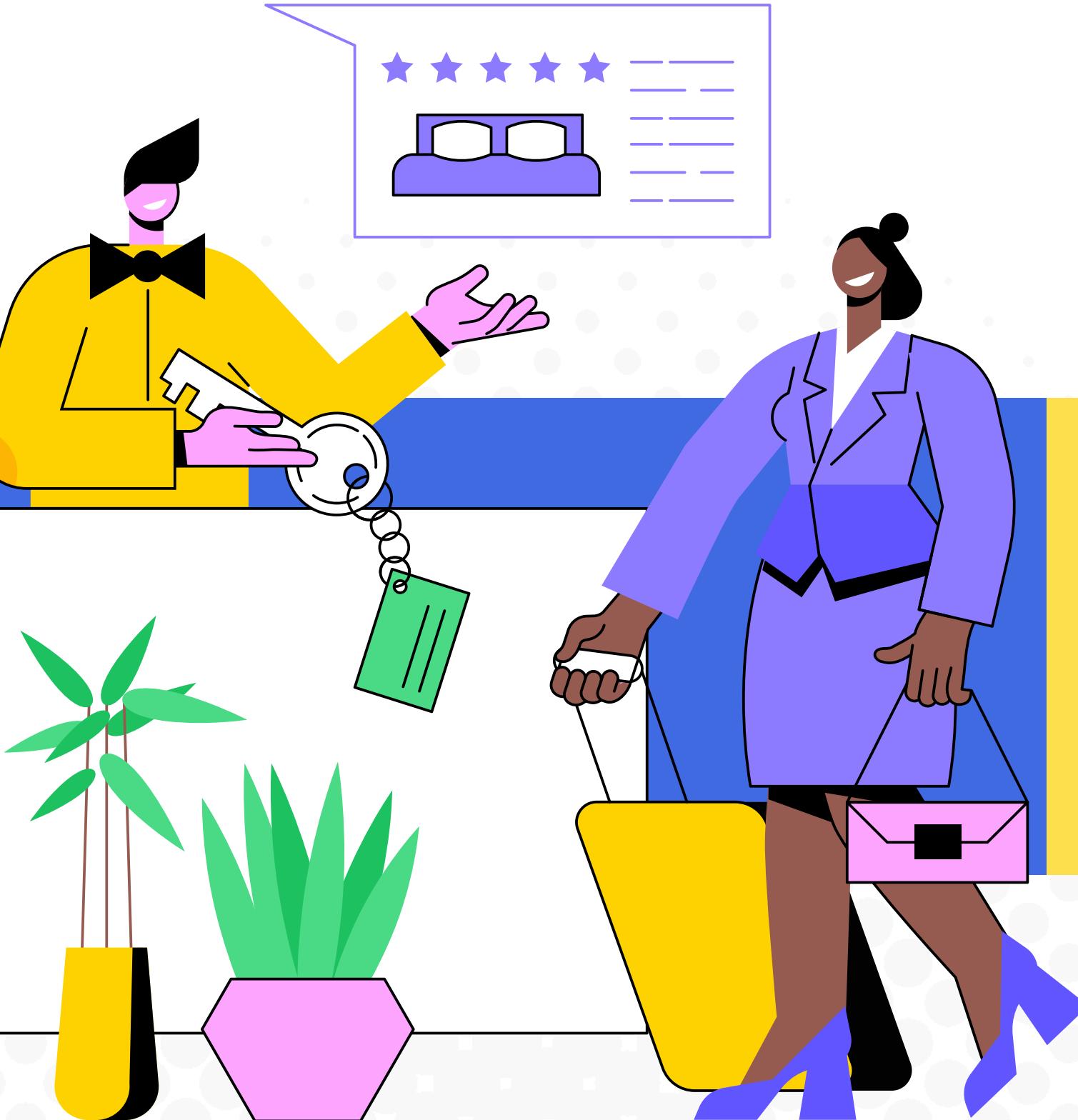
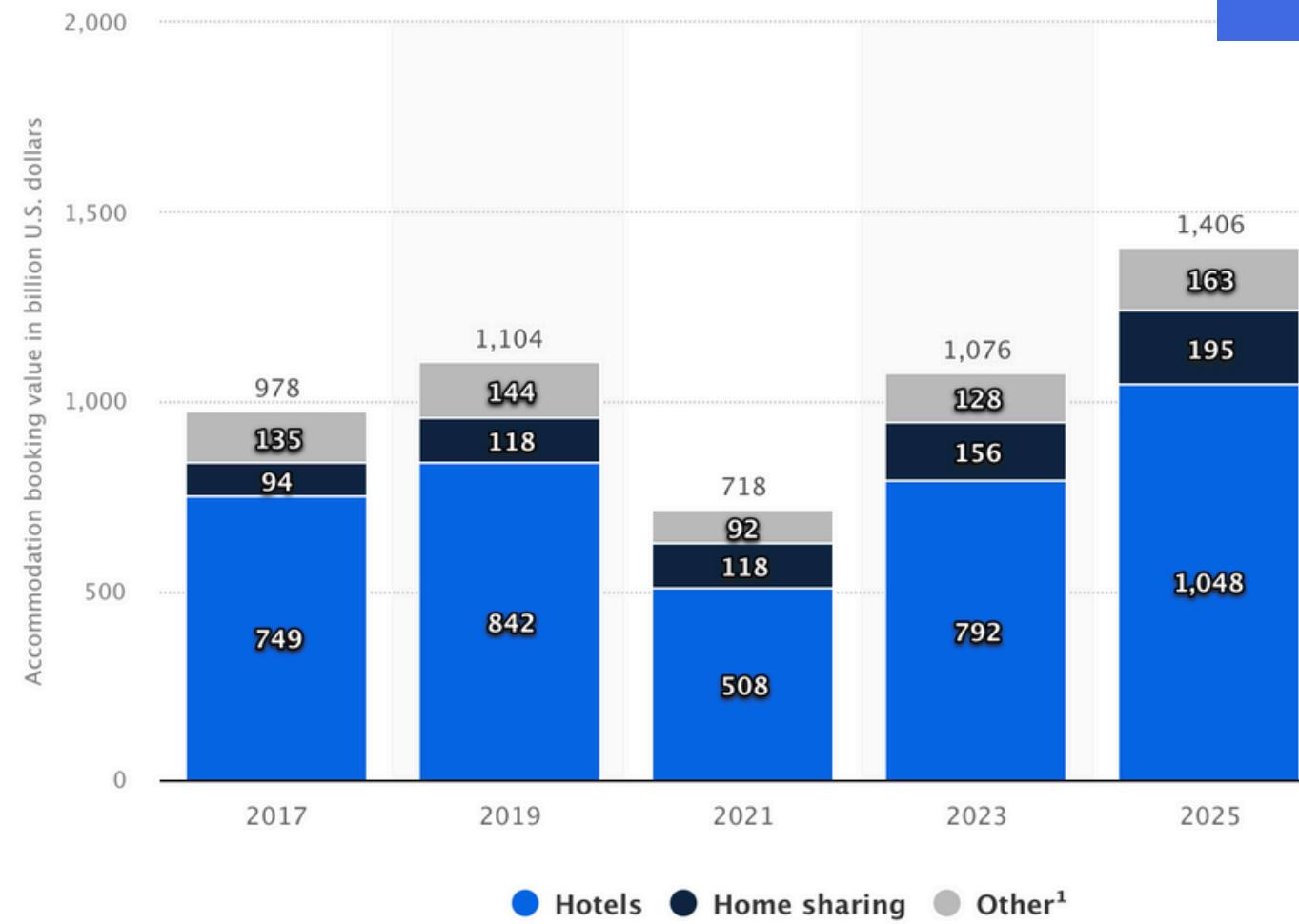
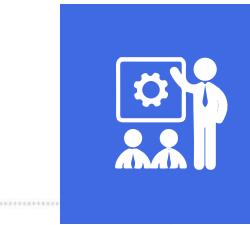




# A machine learning approach for Hotel Booking Cancellations: from data mining to real-time predictions

PROJECT PRESENTATION





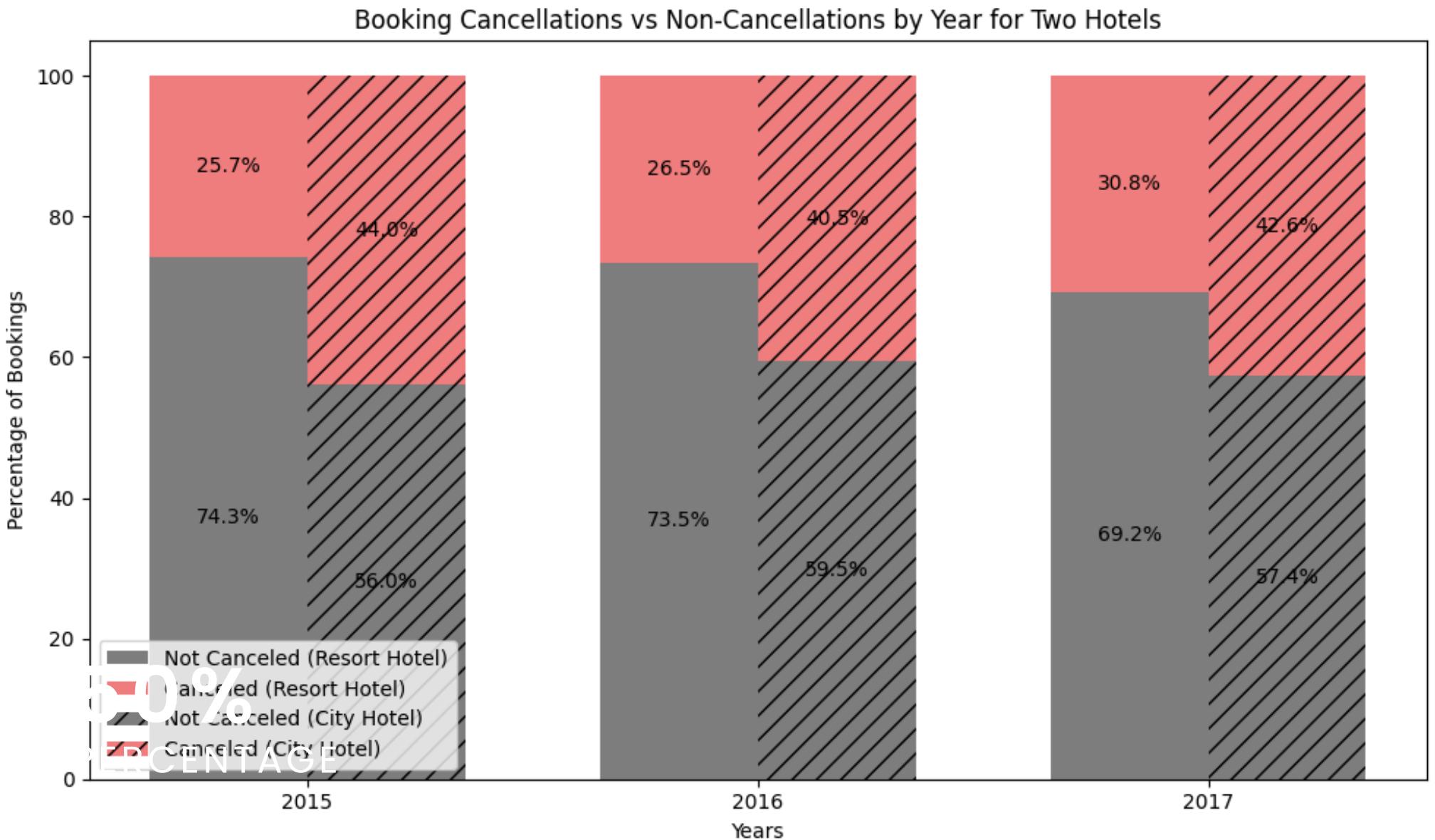
# Introduction

- Booking cancellations significantly impact the hospitality industry, affecting revenue and operations.
  - Predicting cancellations allows hotels to optimize inventory, reduce overbooking, and minimize revenue loss.
  - Machine learning models offer an effective way to predict cancellation probabilities and act proactively.

# Dataset

## Dataset Overview

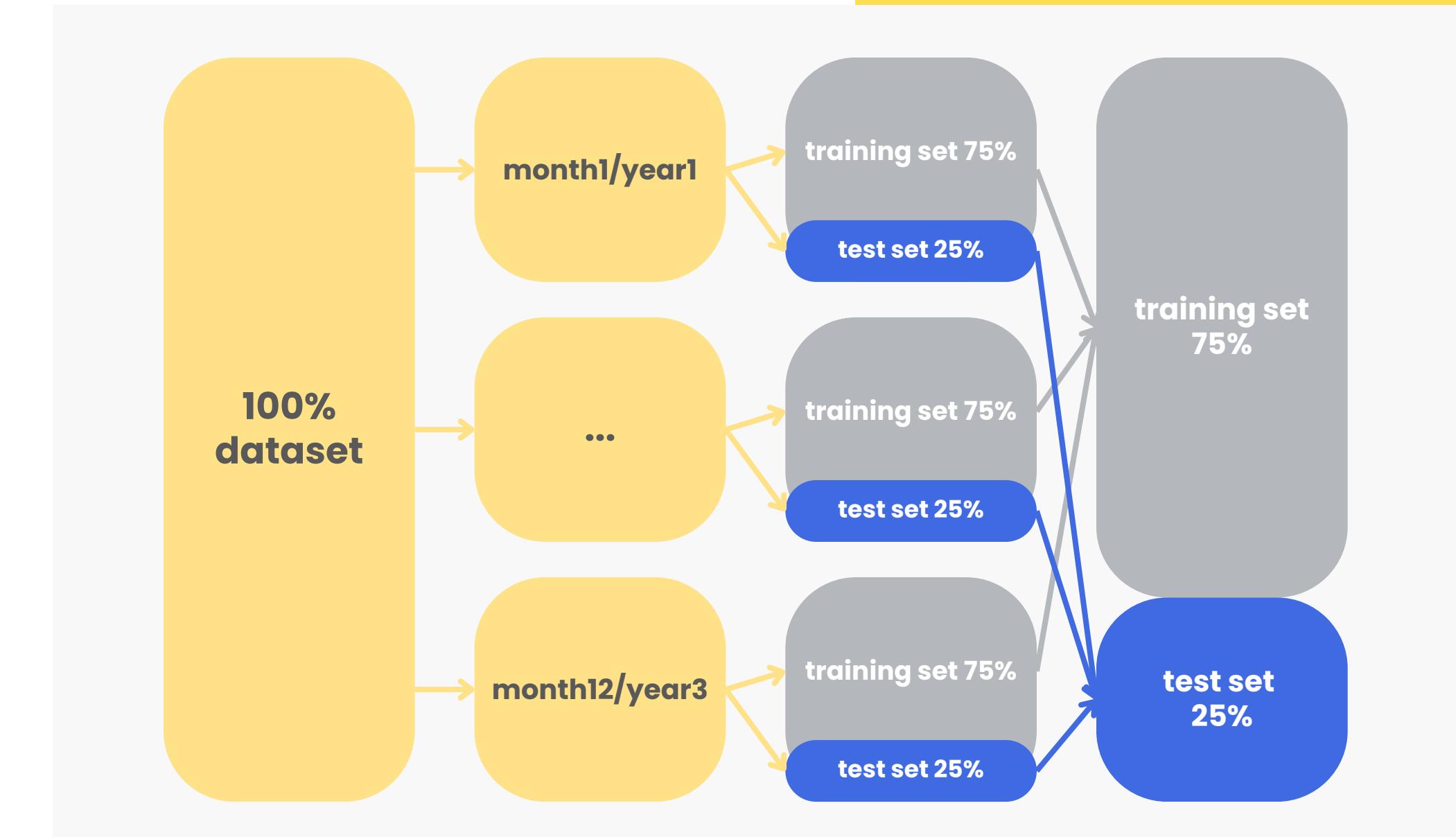
This project utilizes two hotel booking datasets, each containing records from distinct types of hotels: a **resort hotel (H1)** and a **city hotel (H2)**. The data spans a period between July 2015 and August 2017, encompassing all bookings made during this time frame, whether they resulted in cancellations or actual stays.



# Dataset

## DATASET CONVENIENCE SPLITTING

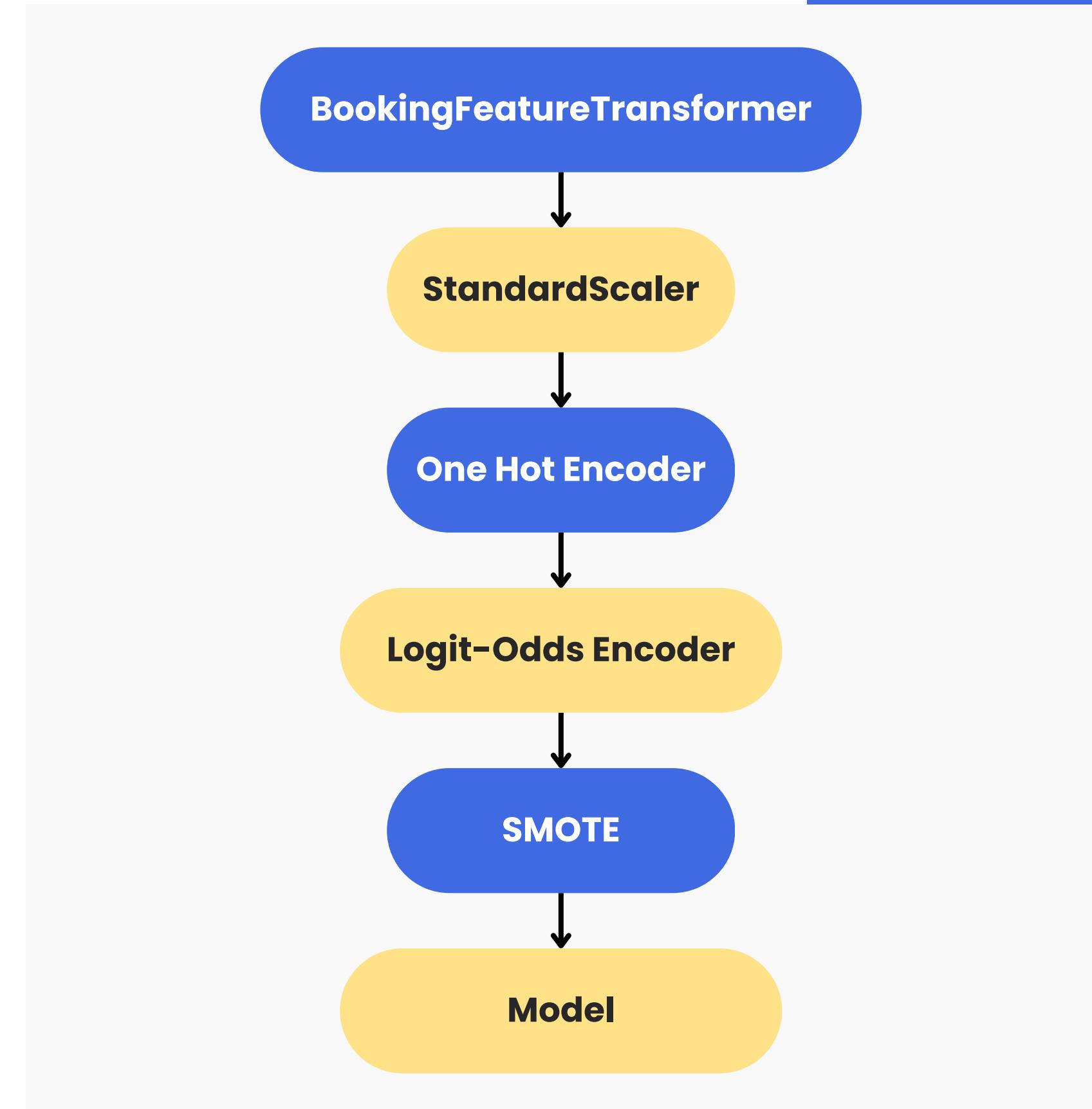
- Time-ordered dataset splitting: **75% training, 25% testing** based on **month-year blocks**.
  - Preserves time structure but focuses on predicting cancellations, not time series.
  - It ensures the model is trained on training data and **evaluated on unseen data**, providing a **more realistic performance assessment**



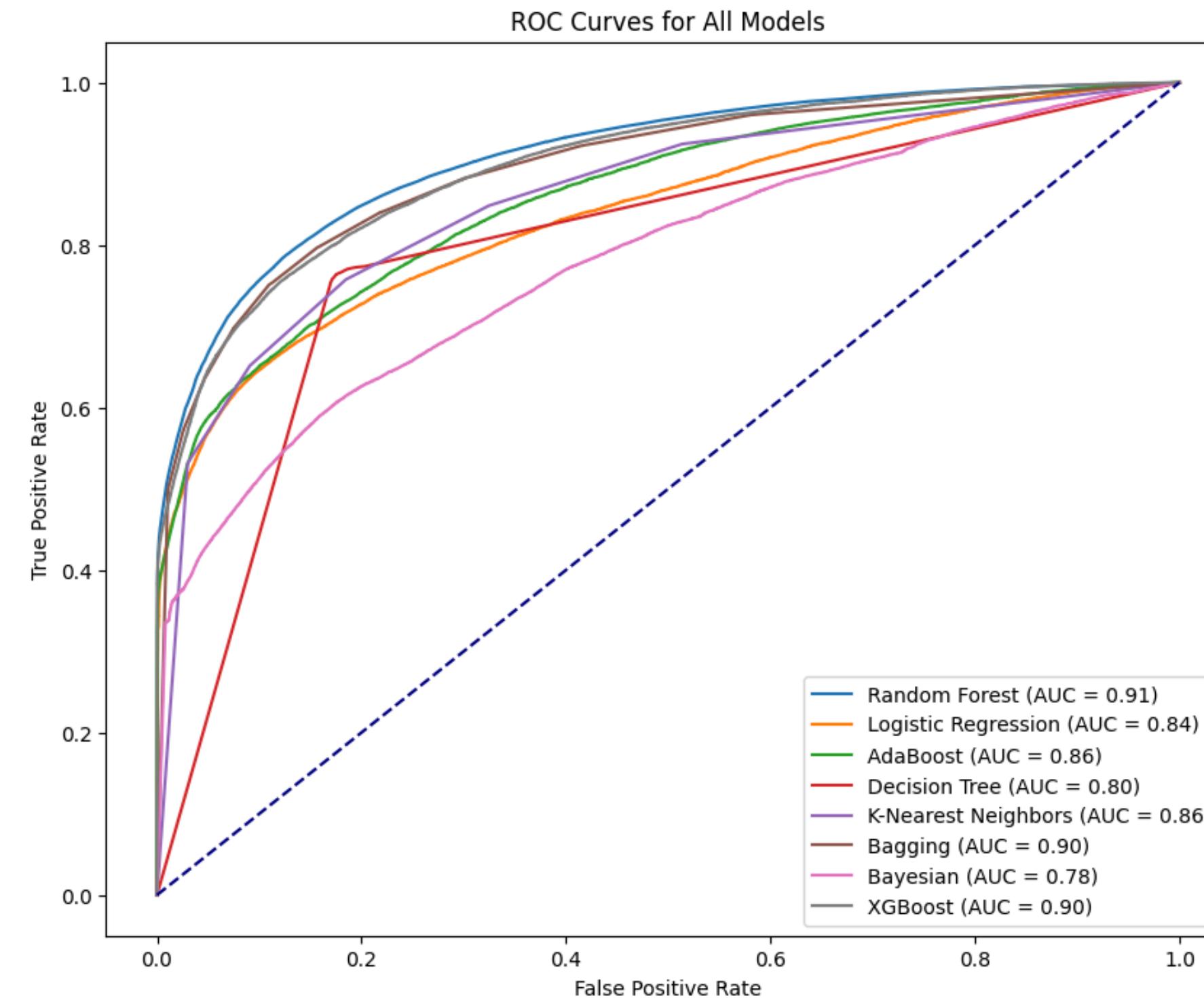
**N.B:** the problem remains a **classification problem**, not time-series forecasting

# Data Preprocessing and Pipeline Build

- **Data Cleaning:** Handled missing values, converted month names to numbers, stripped string labels, dropped noisy and useless columns.
- **Feature Engineering:** Created *ADRThirdQuartileDeviation* to capture ADR variability.
- **High Cardinality Categorical Encoding:** Applied *Logit-Odds Encoding* to Agent and Company features to reduce overfitting

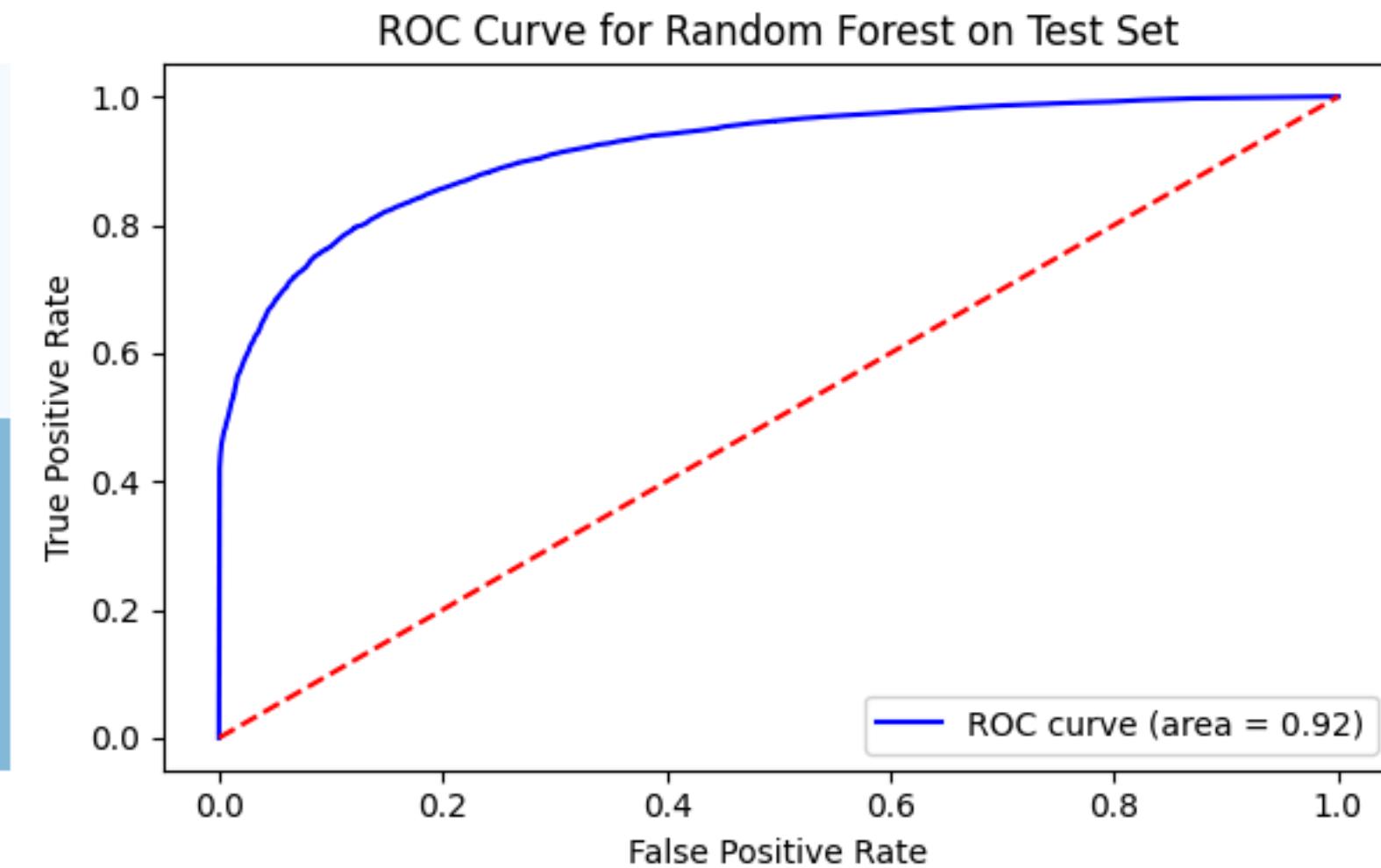
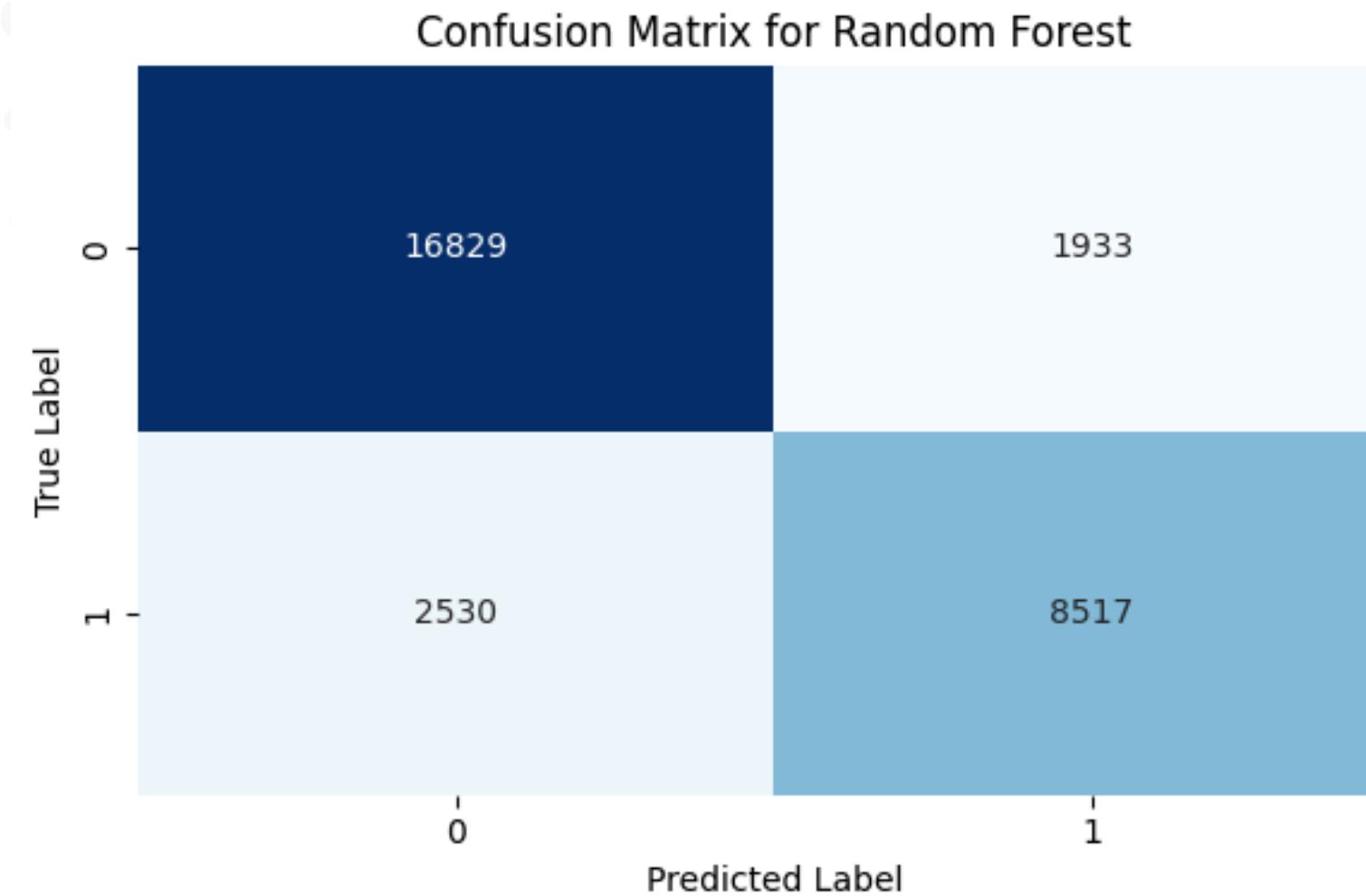


# Models Comparison



		Accuracy	Precision	Recall	F1	ROC AUC
	Random Forest	0.8457	0.8095	0.7635	0.7858	0.9110
	Logistic Regression	0.7855	0.7153	0.7001	0.7076	0.8397
	AdaBoost	0.7924	0.7262	0.7063	0.7161	0.8607
	Decision Tree	0.8022	0.7198	0.7639	0.7412	0.7962
	K-NN	0.7939	0.7073	0.7576	0.7316	0.8580
	Bagging	0.8379	0.7982	0.7531	0.7750	0.8953
	Bayesian	0.6578	0.5260	0.7765	0.6272	0.7787
	XGBoost	0.8348	0.7958	0.7456	0.7699	0.9000

# Model Evaluation



	Accuracy	Precision	Recall	F1
Random Forest	0.8503	0.8150	0.7710	0.7924



# Graphical User Interface

The screenshot displays a graphical user interface for a "Hotel Booking Cancellations Classifier".

**Main Window (Left):**

- Title:** Hotel Booking Cancellations Classifier
- Welcome Message:** Welcome to the Hotel Booking Cancellations Classifier!
- Instructions:** Please enter the following information to predict whether a booking will be cancelled:
- Input Fields (Left):**
  - Lead Time: 180
  - Arrival Year: 2017
  - Arrival Month (1-12): 1
  - Arrival Day of Month: 1
  - Arrival Week Number: 1
  - Stays in Weekend Nights: 1
  - Stays in Week Nights: 2
  - Number of Adults: 1
  - Number of Children: 0
  - Previous Cancellations: 3
  - Is Repeated Guest (1=Yes, 0=No): 0
  - Average Daily Rate: 100
  - Meal: BB
- Input Fields (Right):**
  - Meal: BB
  - Market Segment: Direct
  - Distribution Channel: TA/TO
  - Reserved Room Type: A
  - Deposit Type: No Deposit
  - Customer Type: Transient
  - Agent: 9
  - Company: 40
  - Previous Bookings Not Canceled: 0
  - Total of Special Requests: 0
  - Days in Waiting List: 0
  - Booking Changes: 0
  - Hotel Type: H1
- Predict Button:** Predict

**Modal Dialog (Center):**

- Icon:** A blue folder icon.
- Text:** The predicted class is: Cancelled
- OK Button:** A blue button labeled "OK".



# References

- N. Antonio, A. de Almeida and L. Nunes, "Predicting Hotel Bookings Cancellation with a Machine Learning Classification Model," 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 2017, pp. 1049–1054, doi: 10.1109/ICMLA.2017.00-11.
- Z. A. Andriawan et al., "Prediction of Hotel Booking Cancellation using CRISP-DM," 2020 4th International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2020, pp. 1–6, doi: 10.1109/ICICoS51170.2020.9299011.

## Dataset Paper

- Nuno Antonio, Ana de Almeida, Luis Nunes, Hotel booking demand datasets, Data in Brief, Volume 22, 2019, Pages 41–49, ISSN 2352-3409, <https://doi.org/10.1016/j.dib.2018.11.126>. (<https://www.sciencedirect.com/science/article/pii/S2352340918315191>)



**Thank You**  
FOR YOUR ATTENTION

