🏨 Hotel Reservation Cancellation Prediction

📌 **Description** Development of an ensemble machine learning model to predict booking cancellations using historical hotel reservation data. The solution addresses revenue loss from cancellations through proactive risk assessment.

💡 **Business Value** Predicts cancellations with high accuracy to enable dynamic overbooking, optimized channel management, and targeted guest segmentation, reducing annual revenue loss estimated at $670K for mid-sized hotels.

📈 **Key Results**

- Achieved **87.5% accuracy** and **86% sensitivity** in cancellation detection
- Identified critical drivers: long lead time (3.2x higher risk), repeat guests (68% lower cancellations), and pricing sensitivity ($50 increase → 16% higher risk)

🌟 **Key Benefits**

- Enables revenue protection through data-driven overbooking
- Improves handling of high-risk OTA bookings (40% cancellation revenue impact)
- Supports pricing and promotion strategies to minimize cancellation risk

🧰 **Tech Stack** `Python`, `pandas`, `scikit-learn`, `xgboost`, `catboost`, `imblearn`

📊 **Methods Used** Voting Classifier, Random Forest, Under-Sampling, PCA, Feature Engineering, RobustScaler

🔗 **Links** Repo

Note: Validated on real-world hotel booking data from two Portuguese hotels, with methodology transferable to global hotel operations.