

HOTEL RESERVATION CANCELLATION PREDICTOR

AUTHORS: JAN ERIK JAKSTEIN, MATTIAS RAHNU
GROUP E7

INTRODUCTION

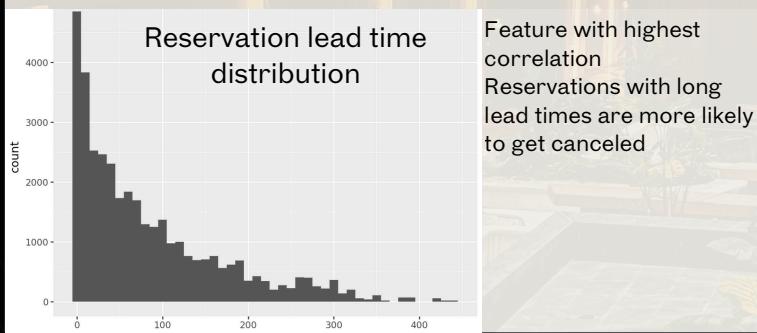
- Hotels lose revenue from bookings that get canceled.
- If hotels can predict when a booking gets cancelled, their revenue loss could be mitigated.

GOALS

- Find the most influential factors that impacts the cancellation rate.
- Develop a roadmap for creating a predictive model for this type of problem.

DATA

- Kaggle dataset
- Two Portuguese hotels, 2017-2018
- Tabular data - 36,275 rows and 19 columns

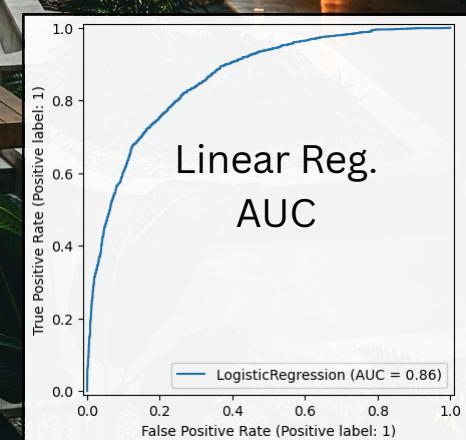
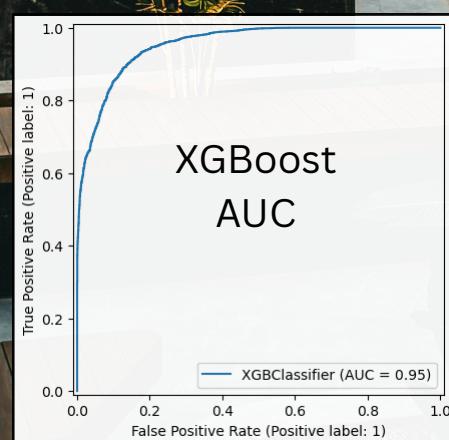
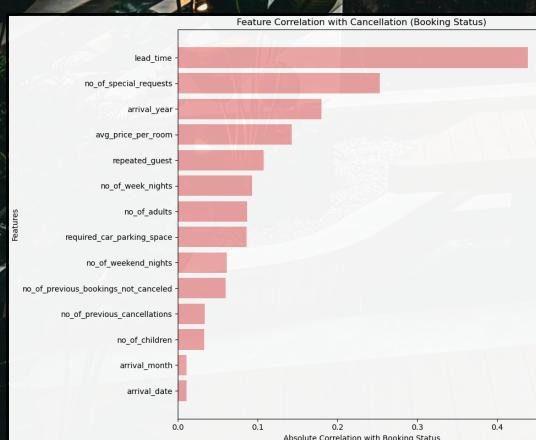


APPROACH

- Data exploration - finding correlating factors to cancellation (distributions, feature engineering, heatmap)
- Model selection and training (train/test split 75%-25%):
 - Logistic Regression
 - Random Forest
 - XGBoost
- Test and find the best model for this data (ROC AUC, Recall)

RESULTS

- Best performing - XGBoost
 - ROC AUC: 0.95
 - Recall: 0.81
- Most influential features:
 - lead_time, no_of_special_requests, avg_price_per_room and market_segment_type
- Successful roadmap for hotels to follow



SOURCES

- Kaggle dataset - <https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset>
- Repository - https://github.com/Jantz/ids-2025-hotel_reservations