

oooo



HOTEL BOOKING CANCELLATIONS PREDICTION USING MACHINE LEARNING

oooo

TABLE OF CONTENTS

- Introduction
- Objective
- Exploratory Analysis
- Model Summaries
- Conclusions



INTRODUCTION

The need for lodging is increasing day by day. This is related to the number of trips made far from their place of residence. Usually, before booking lodging, they will make a reservation. Nowadays, people who are just planning a trip can already make sure they have a place to stay before they even leave their homes. This kind of reservation is also applied to several requirements that must be met. Even if specific requirements have been met, there are often many unforeseen conditions that cause a reservation to be forced to be canceled.

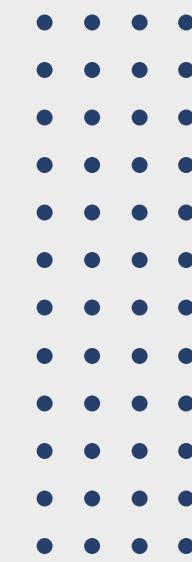
Performing and comparing cancellation prediction using the decision tree and ensemble machine learning model.

OBJECTIVE

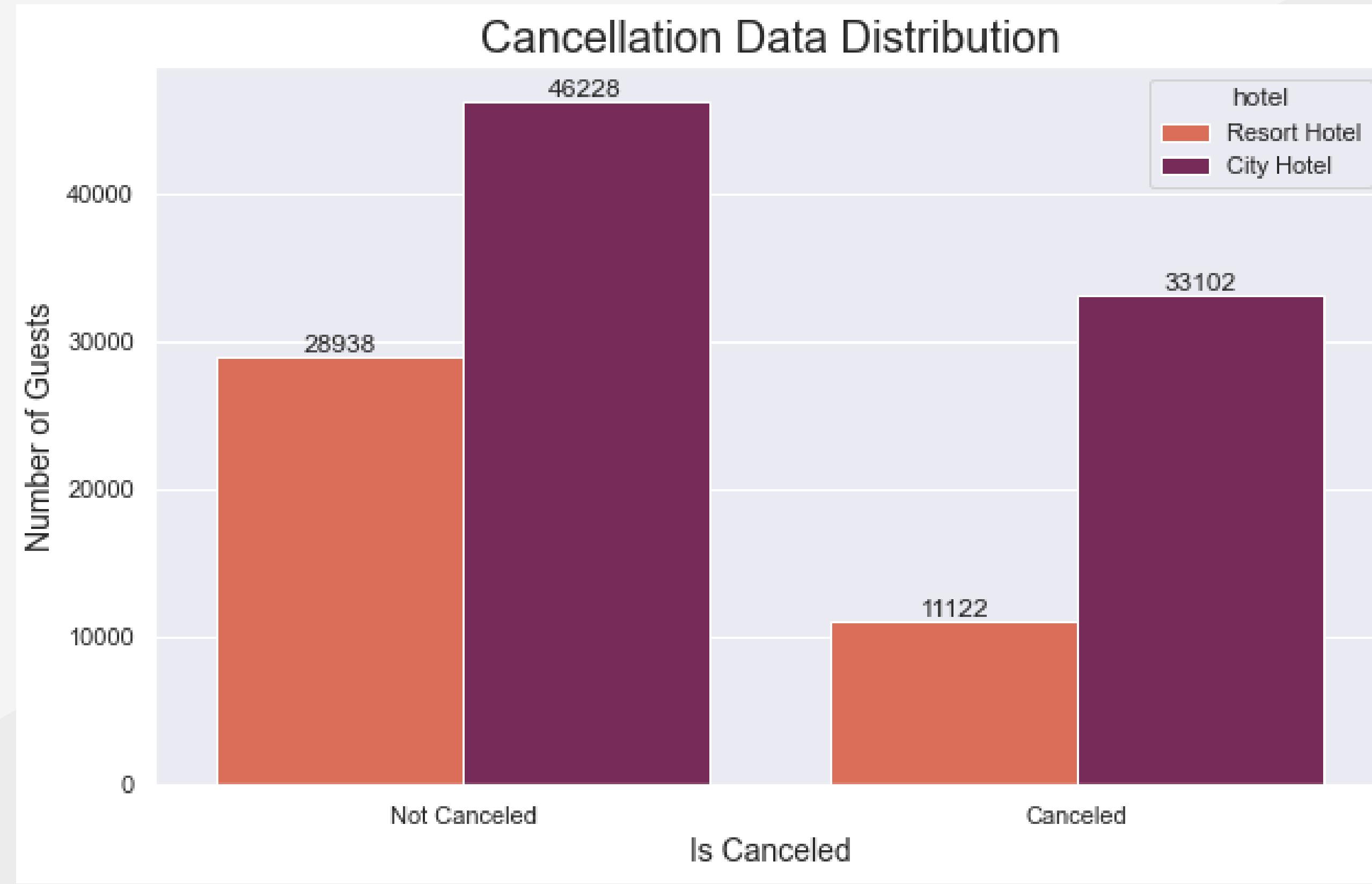
Performing improvement for machine learning model with the best performance.

EXPLORATORY ANALYSIS

ooo



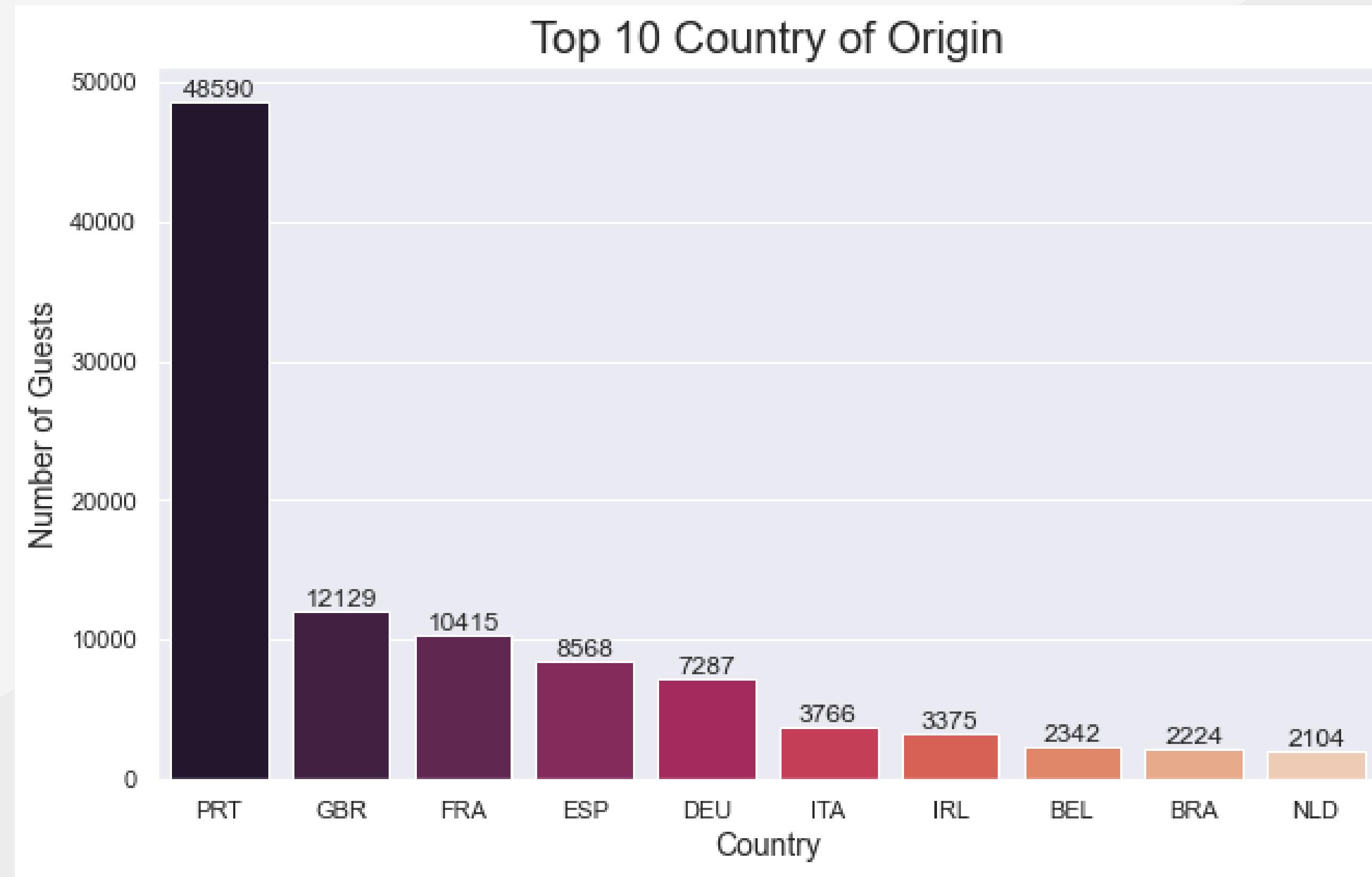
Cancellation Data Distribution



EXPLORATORY ANALYSIS

0 0 0 0

Top 10 Country of Origin



EXPLORATORY ANALYSIS

OOOO

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

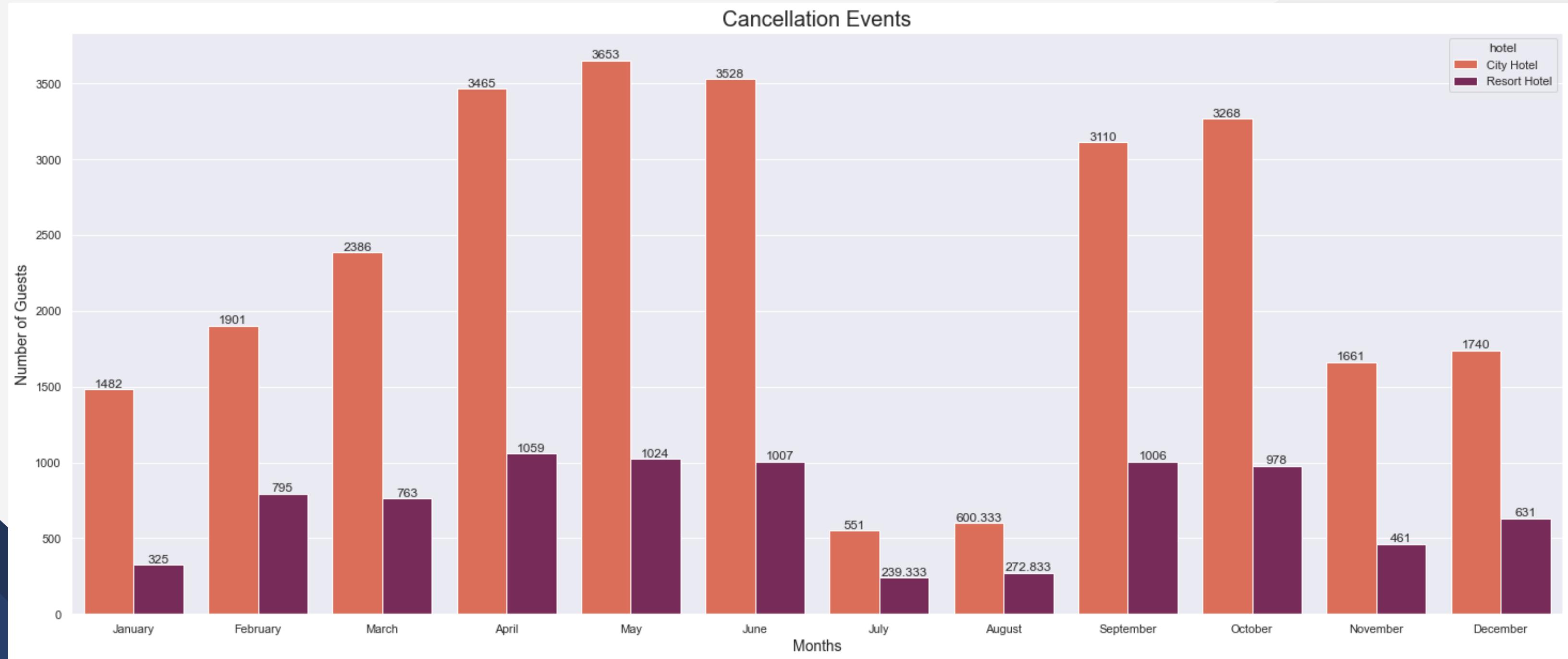
•

•

•

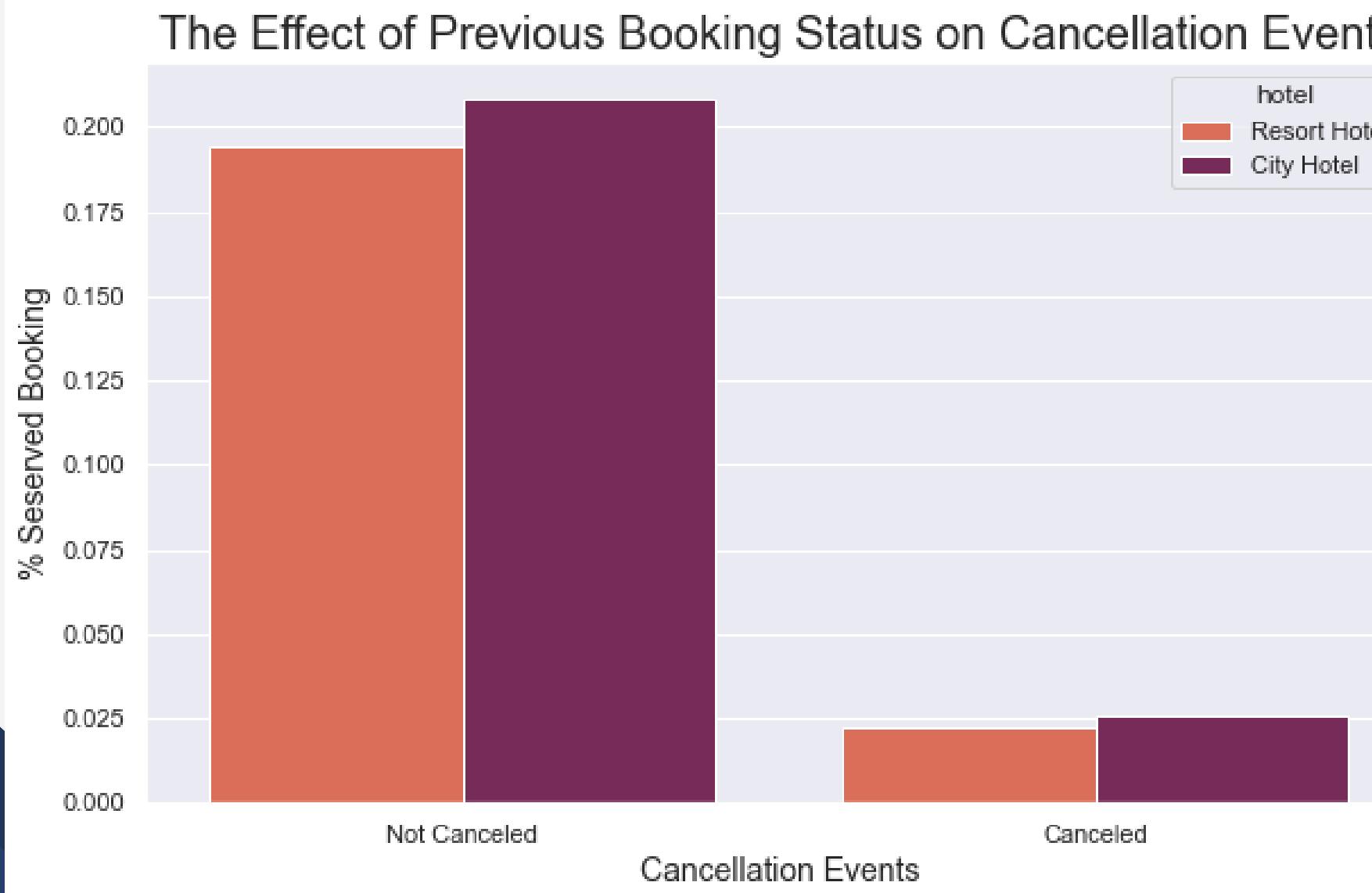
•

•



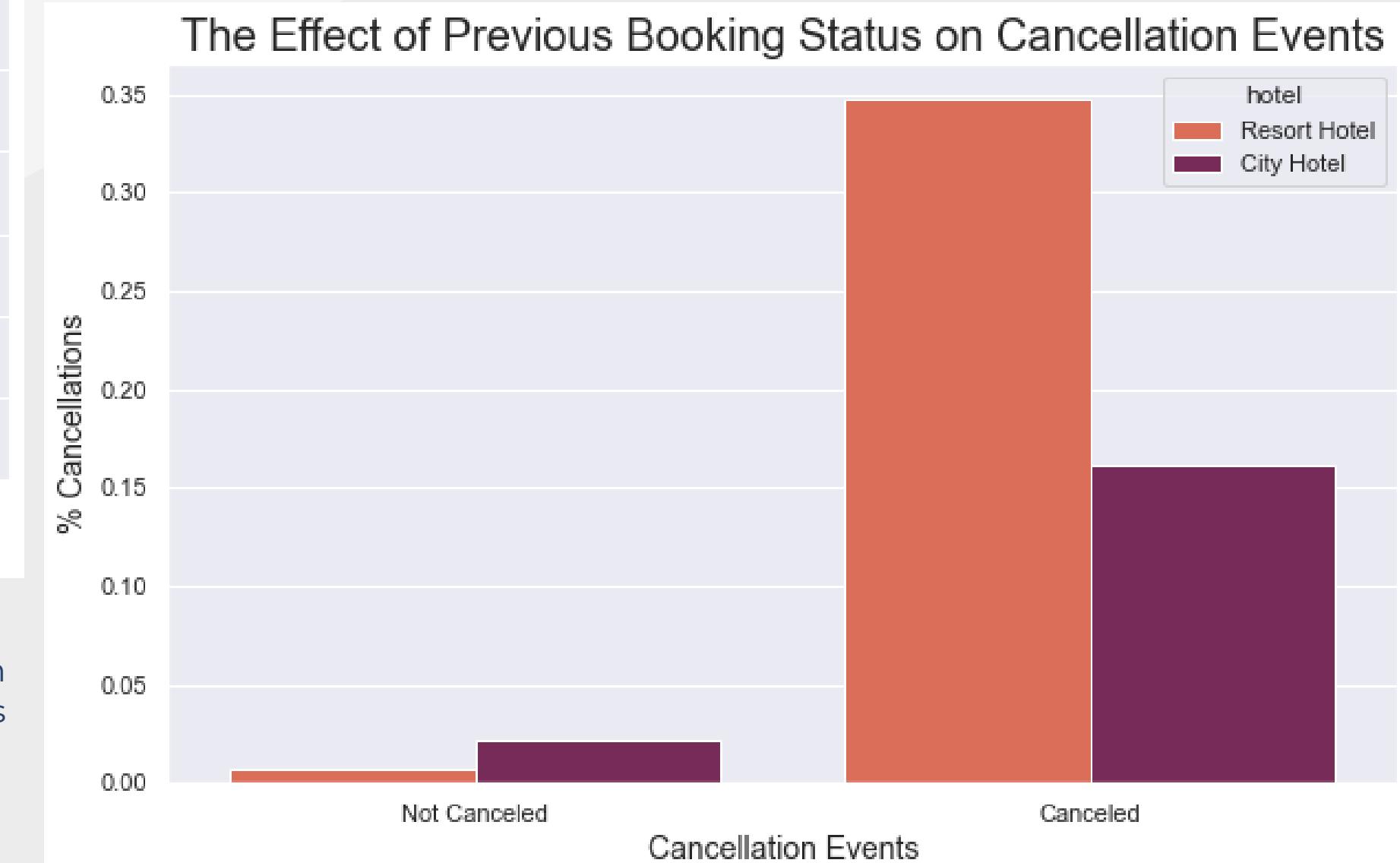
EXPLORATORY ANALYSIS

OOOO



Cancellation events

Reservation events



Cancellation Events

MODEL SUMMARIES

Metric	Decision Tree	Random Forest	Gradient Boosting	XGBoost
Accuracy Train	0.99	0.99	0.73	0.72
Accuracy Test	0.71	0.74	0.67 (70)	0.67

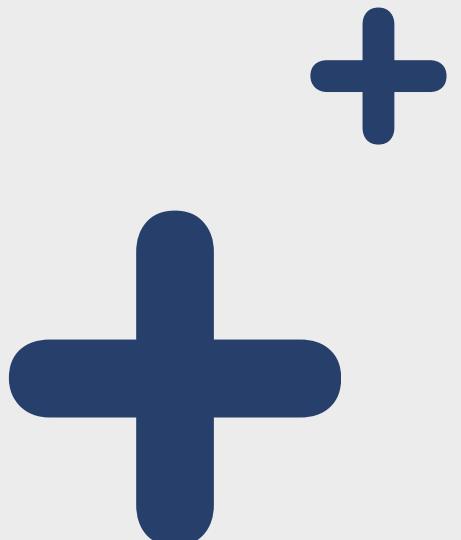
CONCLUSIONS

Based on the modeling results, the decision tree and random forest are the models with the worst performance. Although the trained model has high accuracy, the test model has very low accuracy. Overfitting on these two models occurs.

The other two models which are sequential ensemble learning based, namely gradient boosting and XGBoost have better performance. Although the accuracy of the model is low, there are no overfittings.



The improved model is carried out using random search hyperparameter tuning to the gradient boosting model. The accuracy of the testing model has increased from 0.67 to 0.70.





**THANK
YOU**

