Business Analytics Project Management

Model Evaluation and Recommendations: Predicting Hotel Cancellations

---

Presented by:

Hojin Lee • Qing Liu • Mostafa Salem • Shristi Lamsal • Sky Song • Yana Olshevskaya • Yu Zhang

Project Objective

Predicting customers who are likely to cancel their hotel reservations and providing recommendations that will allow hotels to reduce the cancellation rates.

09/27
Project proposal

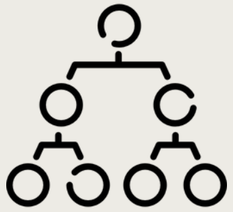
10/25
Data understanding and preparation

11/15
Model selection

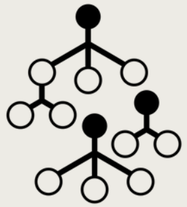
12/13
Model evaluation and recommendations

Predictive Analytics - Models

Decision Tree

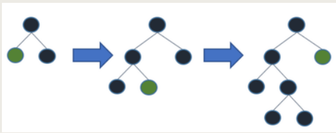


Random Forest



Logistic Regression

LGBM (Light Gradient Boosting Machine)



Model Evaluation Metrics - Minimizing false negatives

1. **Confusion Matrix** - A table that displays and compares actual values with the model's predicted values.
2. **Accuracy** - The percentage of correct decisions.
3. **Precision** - How many predicted positives are actually correct.
4. **Recall** - The proportion of the positive class that got correctly classified.
5. **F1 Score** - The balance between Precision and Recall; deals with the uneven class distribution.
6. **Gain and lift charts** - A visual way to evaluate the effectiveness of different models.
7. **ROC (Receiver Operating Characteristics)** - A probability curve.
8. **AUC (Area Under The Curve)** - The degree or measure of separability. It tells how much the model is capable of distinguishing between classes.

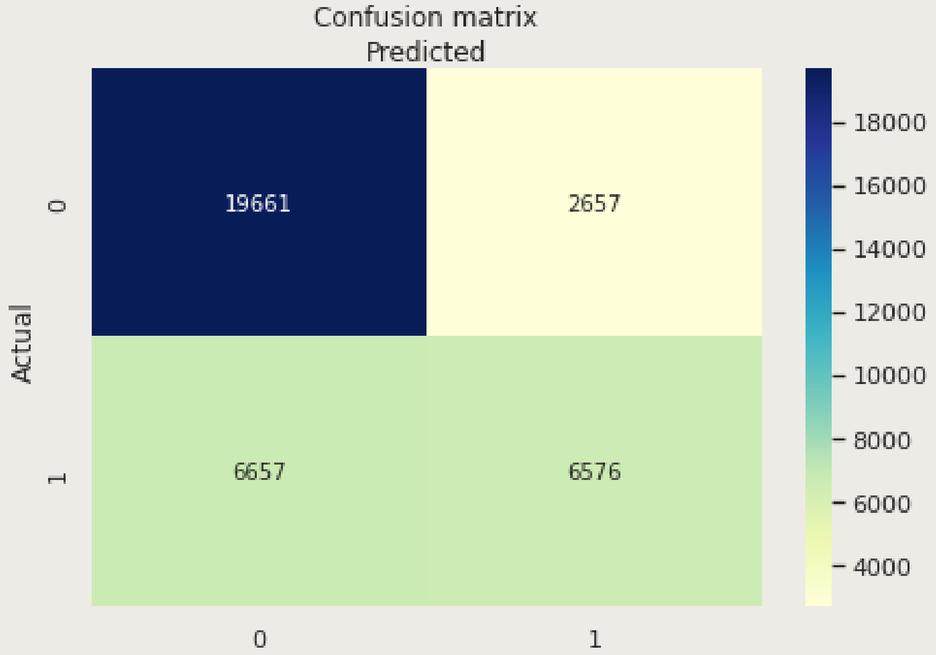
Model Evaluation - Logistic Regression

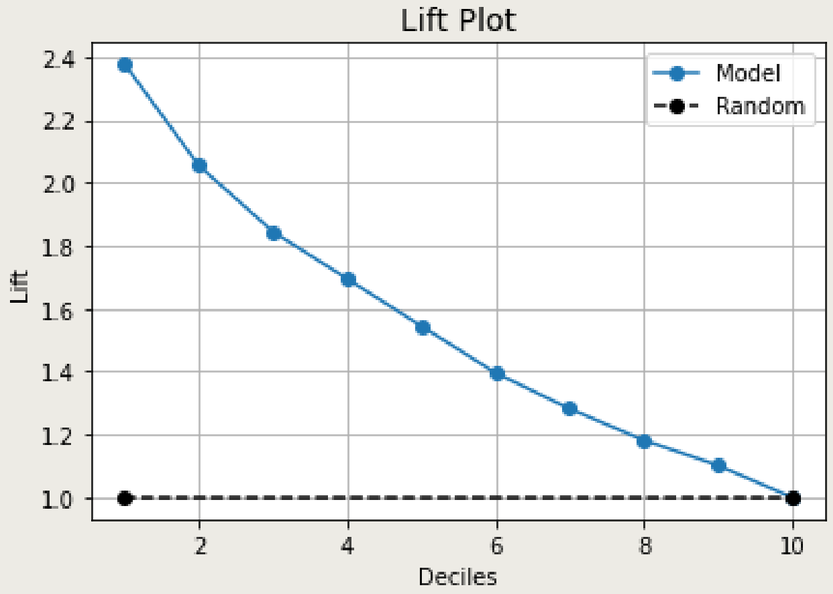
Accuracy: 0.738

F1: 0.585

Precision: 0.712

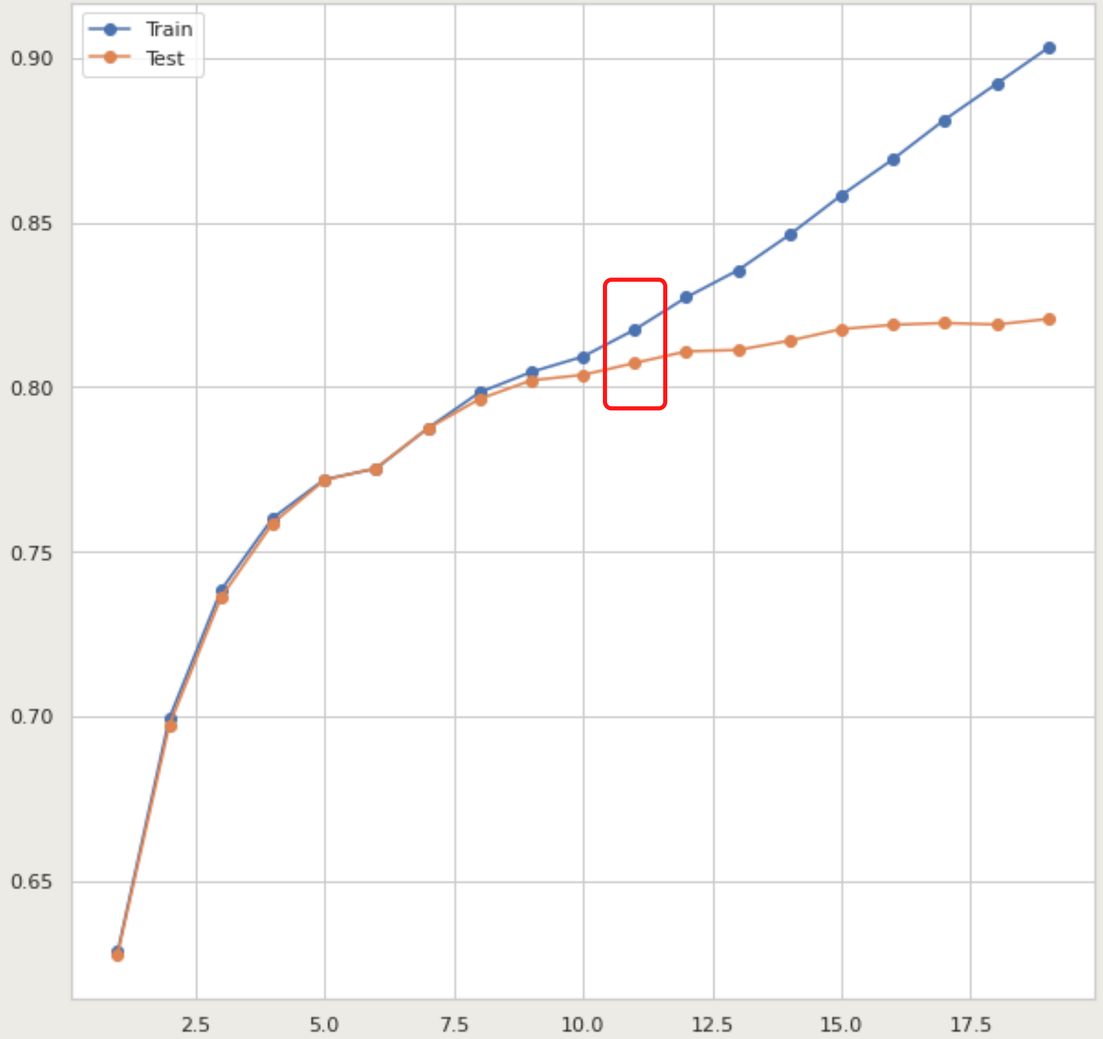
Recall: 0.496





Model Evaluation - Decision Tree

Set decision tree depth to 11



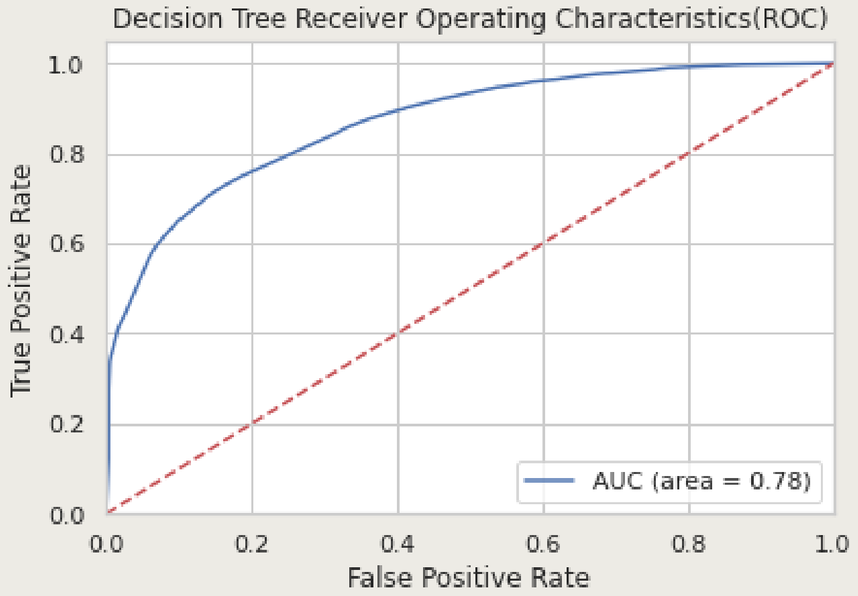
Model Evaluation - Decision Tree

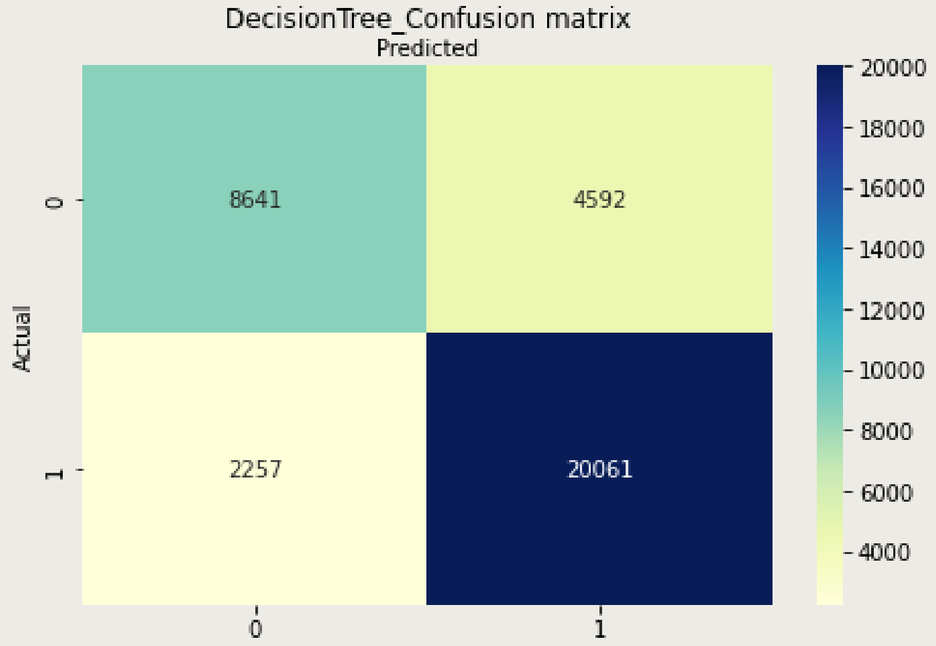
Accuracy: 0.807

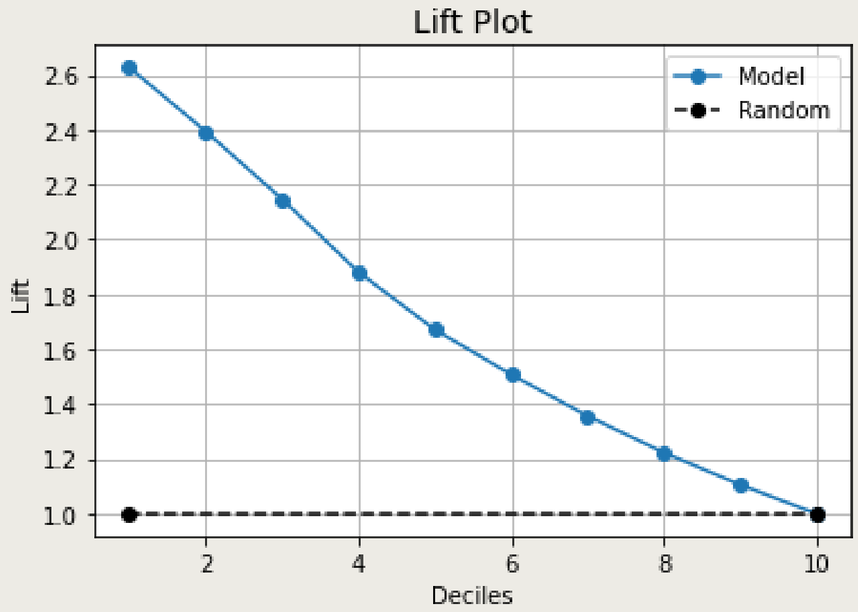
F1: 0.716

Precision: 0.793

Recall: 0.653







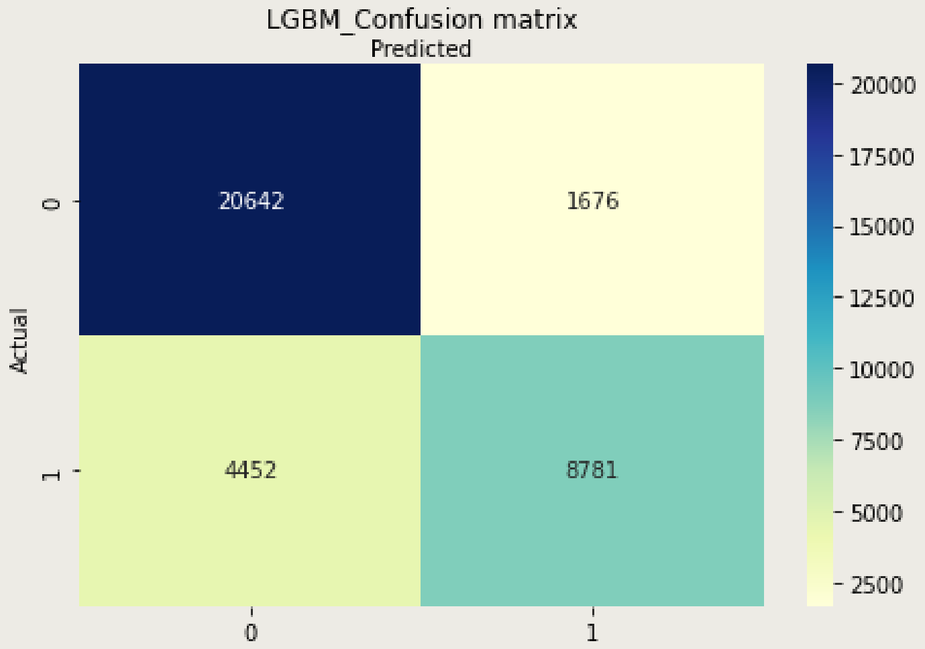
Model Evaluation - LGBM
(Light Gradient Boosting Machine)

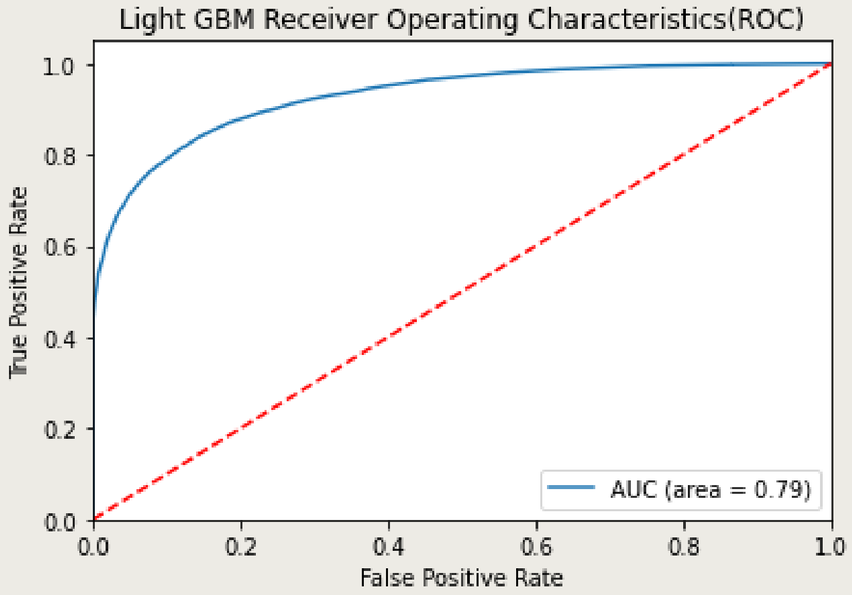
Accuracy: 0.830

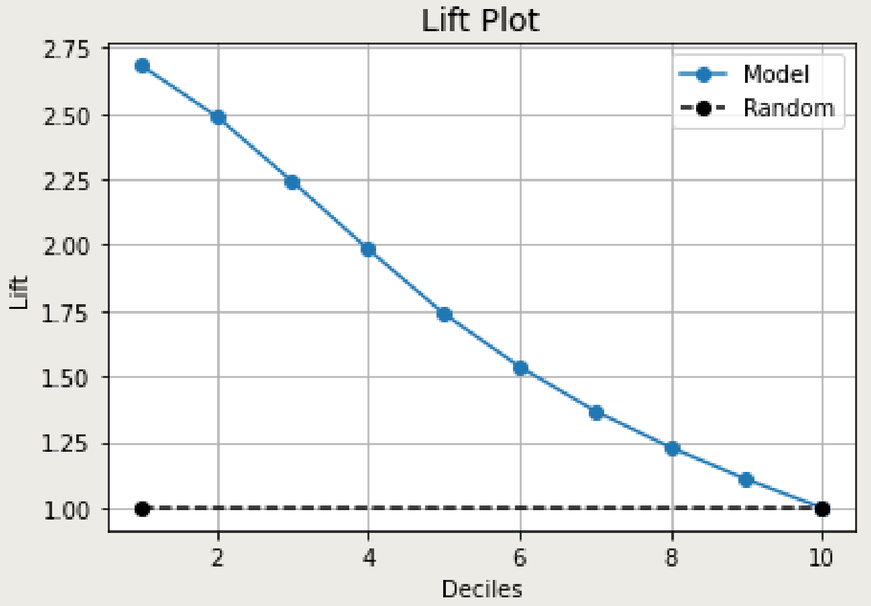
F1: 0.760

Precision: 0.801

Recall: 0.723







Model Evaluation - Random Forest

Overfitting | SMOTE

SMOTE (Synthetic Minority Over-sampling Technique) - statistical technique for increasing the number of cases in a dataset in a balanced way. Reduces overfitting.

Accuracy before SMOTE
Training: 0.991
Test: 0.863

→

Accuracy after SMOTE
Training: 0.944
Test: 0.863

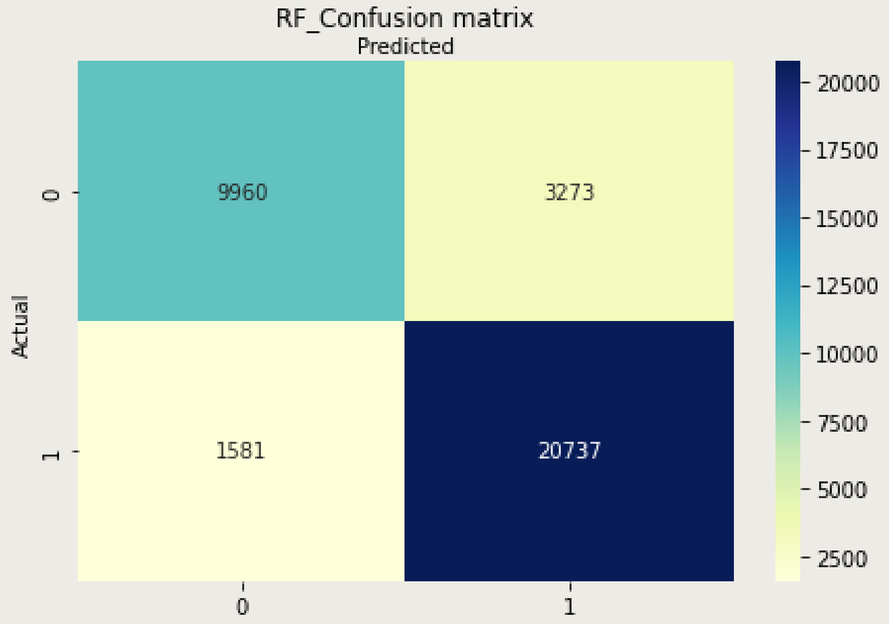
Model Evaluation - Random Forest

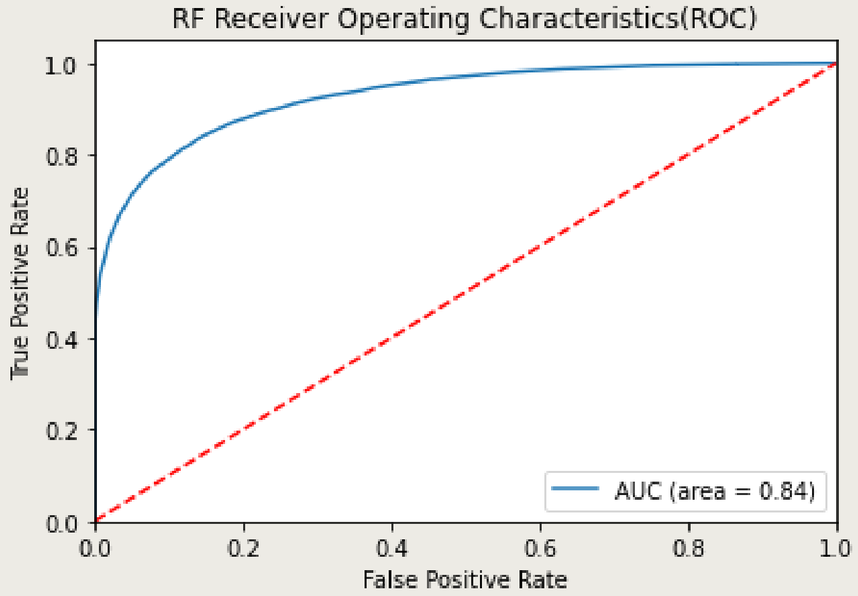
Accuracy: 0.863

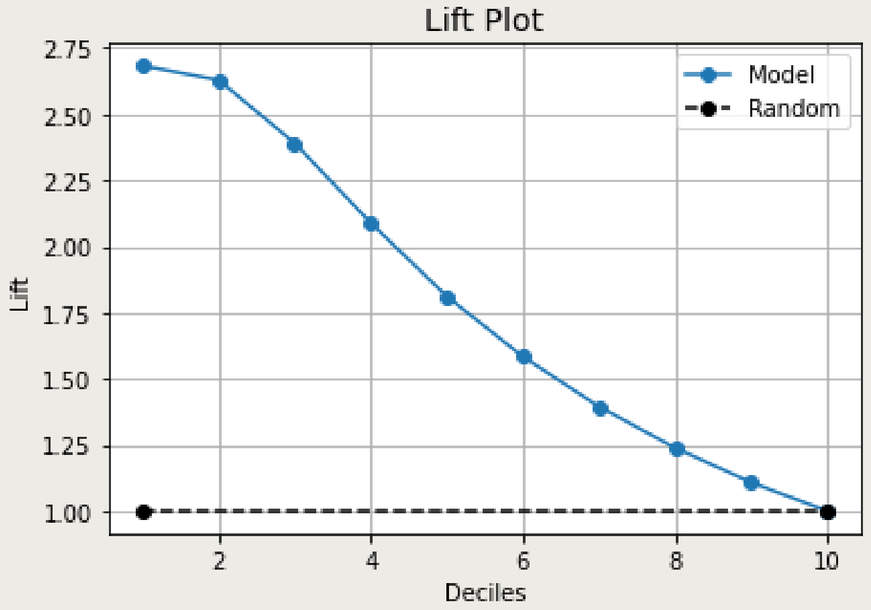
F1: 0.804

Precision: 0.863

Recall: 0.753

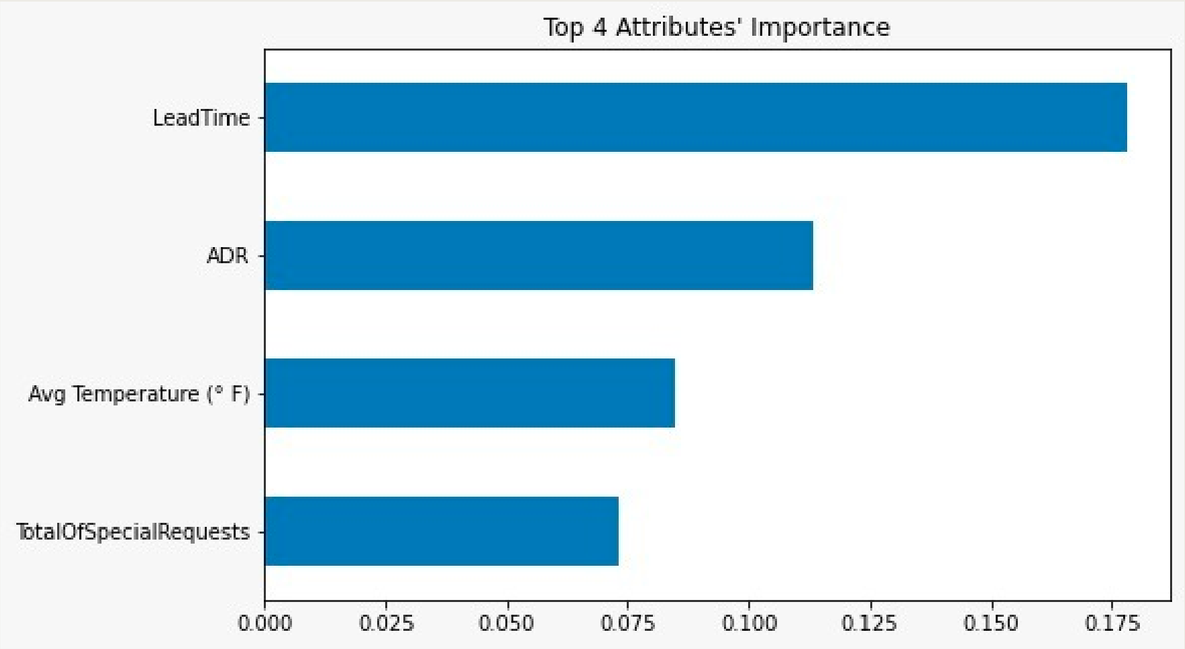




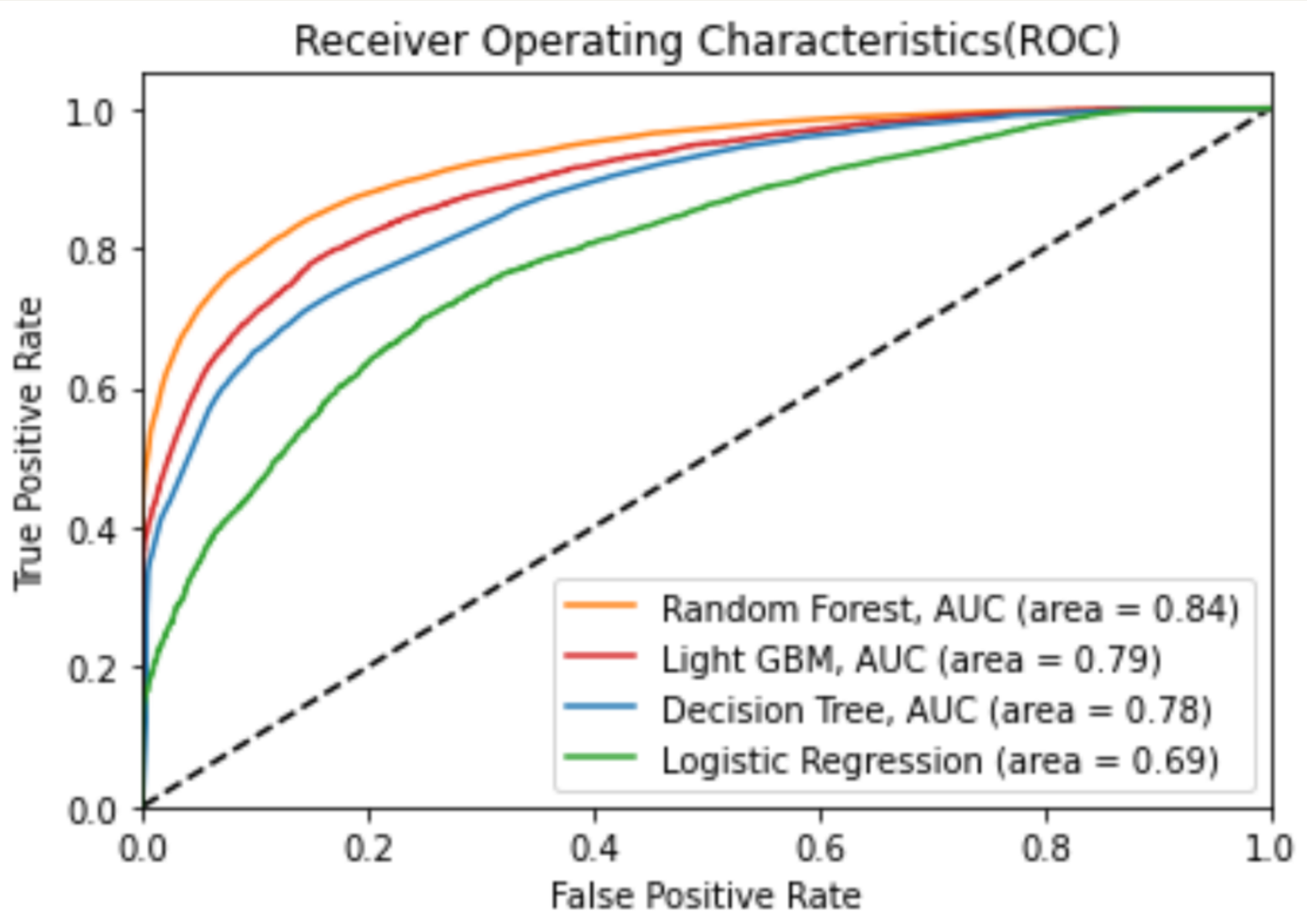


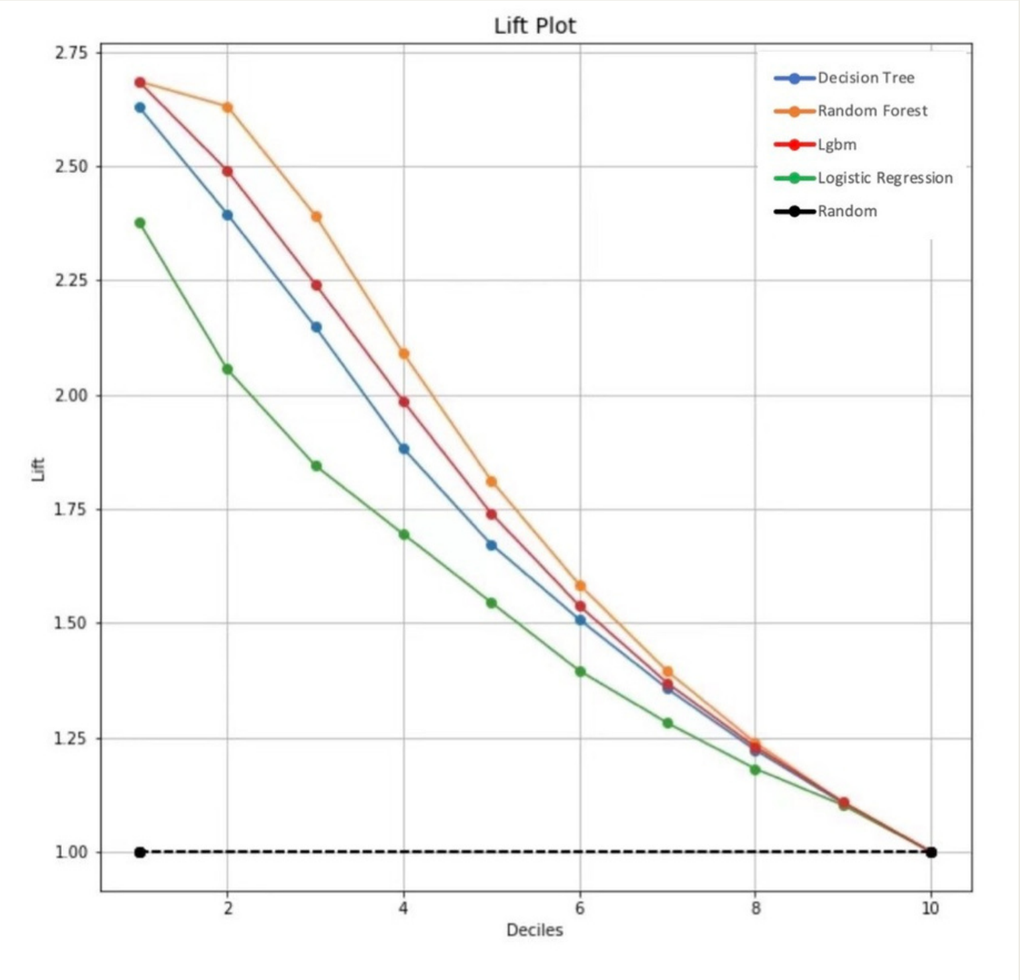
Best Model - Random Forest

Classifier	Accuracy	Precision	Recall	F score
Decision Tree Classifier	0.807460	0.792866	0.653442	0.716434
Random Forest	0.863464	0.863010	0.752664	0.804069
Logistic Regression	0.738010	0.712228	0.496939	0.585418
Light GBM	0.829900	0.800800	0.722800	0.759800



Best Model - Random Forest





Recommendations to Mitigate Cancellations

Project findings:

- Customers who had a special request were 26% less likely to cancel.
- Customers booking 30+ days in advance were 18% more likely to cancel.
- Winter months have the lowest demand and a high cancellation rate.
- Bookings with a lower average daily rate of under 60 EURO have the lowest cancellation rate.

Lead Time

- Discounted rates on non-refundable bookings.
- Personalized emails to customers who are most likely to cancel.
- Email guests one and two weeks before arrival.

Total Special Request

- Offer add-ons to personalize the stay.
- Have an option to declare the occasion for the booking and offer special requests at discounted price.

Weather

- Include information about hotel amenities when sending booking reminders. For example, SPA information when the weather is predicted to be cold.

Average Daily Rate

- Offer discounted bundle packages (hotel + full board + SPA) during winter months.

Q & A