Report to Stakeholders:

Booking Cancellations Insights

What the Models Say about Cancellations?

- **ADR and Lead Time**

Both models showed that these are important features. I am satisfied with the Random Forest (RF) results, which highlight lead time and ADR as the most important features for predicting cancellations, as this makes logical sense. RF does not directly show feature direction or exact importance, but the fact that these variables contribute significantly to tree splits indicates a strong influence on predictions.

We can also assume their relationship with cancellations is positive. First, this aligns with common sense, and second, it matches the corresponding positive odds in Logistic Regression (LogReg). In other words, a high ADR might cause a customer to cancel after finding a cheaper alternative, and a high lead time increases the likelihood of cancellation since plans often change over time.

- **Number of Previous Cancellations**

Both models showed that the number of previous cancellations is also important (5th highest odds ratio in LogReg and 10th most important feature in RF). This is expected, as people who canceled in the past are more likely to cancel again.

- **People from Portugal**

People from Portugal are more likely to cancel, but I cannot interpret this confidently without feedback from the stakeholders. However, I assume the hotels are located in Portugal, as most customers are Portuguese. In that case, there's a possible explanation: natives may be more

flexible knowing the local areas and having more flexible dates, while international travelers usually have fixed travel plans, making them less likely to cancel.

- **Online Travel Agencies**

Both models showed a strong connection between OTA bookings and cancellations (2nd in LogReg odds and 4th in RF importance). We can also assume a positive direction here, based on both the models and logic. One explanation is that the online environment makes booking easier. Customers can book without paying upfront and compare many options at the same time. This might lead to spontaneous bookings, which are more likely to be canceled later.

Suggestions to the Stakeholders

- **ADR & Online Travel Agencies**

Hotels should use Property Management Systems (PMS) with dynamic pricing, uploading real-time price updates to OTAs. Dynamic pricing must adapt to market conditions, seasonal trends, and customer behavior.

A good Machine Learning model can support this in real-time by learning from historical data and ongoing booking trends. It can identify patterns where cancellations are more likely and suggest price changes to reduce risk.

Also, modern PMS systems aim to reduce OTA dependency by increasing direct bookings, which usually bring higher margins and lower cancellation rates.

- **Lead Time**

Hotels could apply motivation strategies like “book for the next month and get a discount or a gift” to encourage short-term bookings and reduce the risks of long lead times. For bookings made far in advance (e.g., more than one year), they could offer a small discount only if the booking is paid upfront. This way, both the hotel and the customer benefit from security and commitment.

- **Number of Previous Cancellations**

A solution here is to create personalized offers for users who canceled in the past. These could be combined with lead-time strategies, such as offering upfront-pay discounts for frequent cancelers, to reduce future risk.

- **Portuguese People**

There is a mutual benefit on supporting the domestic market with discounted direct booking options targeted at Portuguese customers, especially during low-demand seasons. This builds loyalty and may reduce the overall cancellation rate from locals.