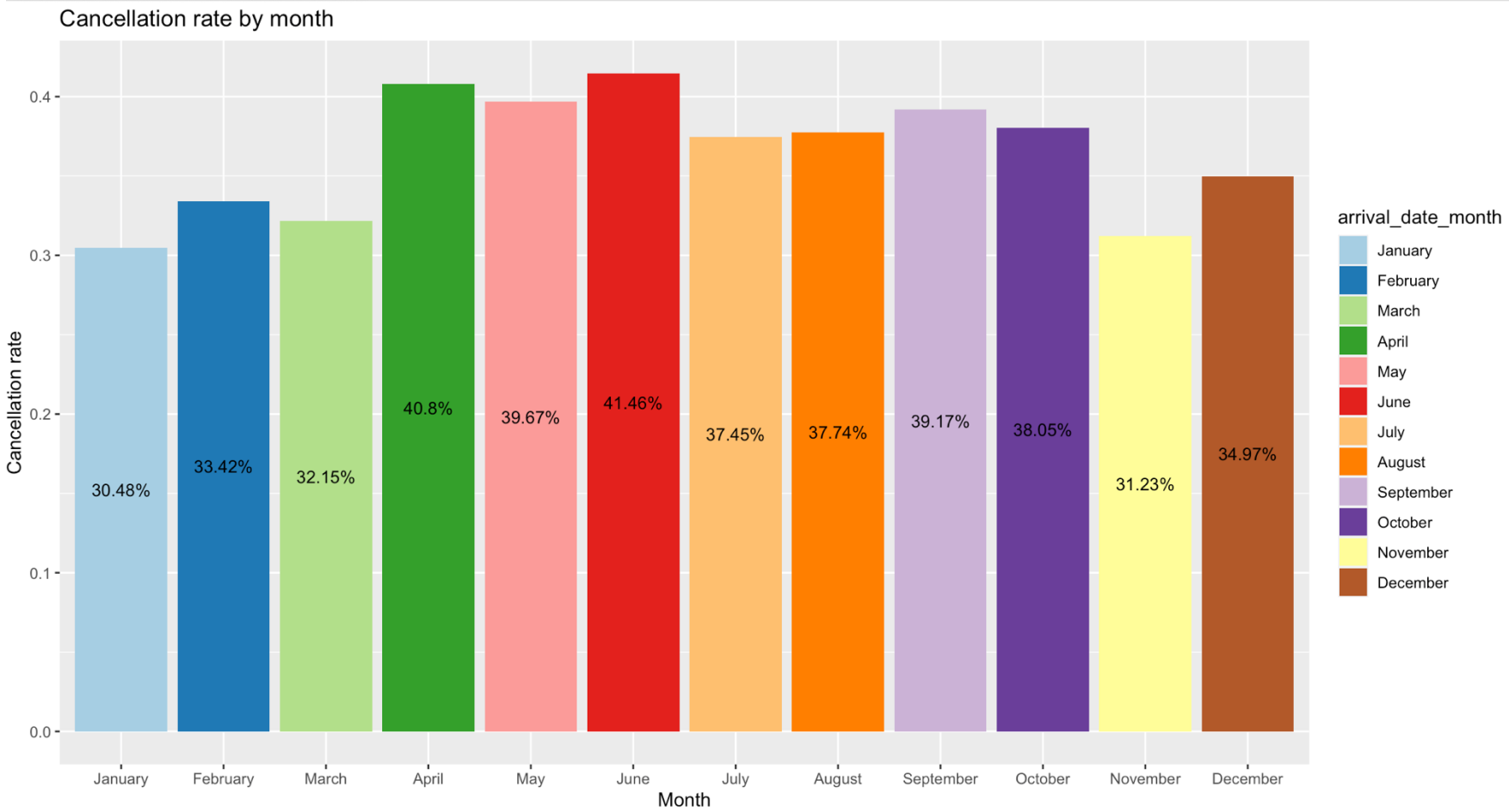
Hotel Cancellation Analysis

Data Source

https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand

- **Time Period:**
  July 01, 2015, to August 29, 2017
- **Hotel Location:**
  Portugal
- **Content:**
  Hotel bookings Dataset focusing on factors influencing reservation cancellations
- **Key Variables:**
  Hotel Type, Is_Canceled, Lead_Time, Arrival_Date_Year/Month/Week_Number/Day_of_Month/ Stays_In_Weekend_/Stays_In_Week_Nights, Adults, etc.

EDA – Cancellation Rate by Month

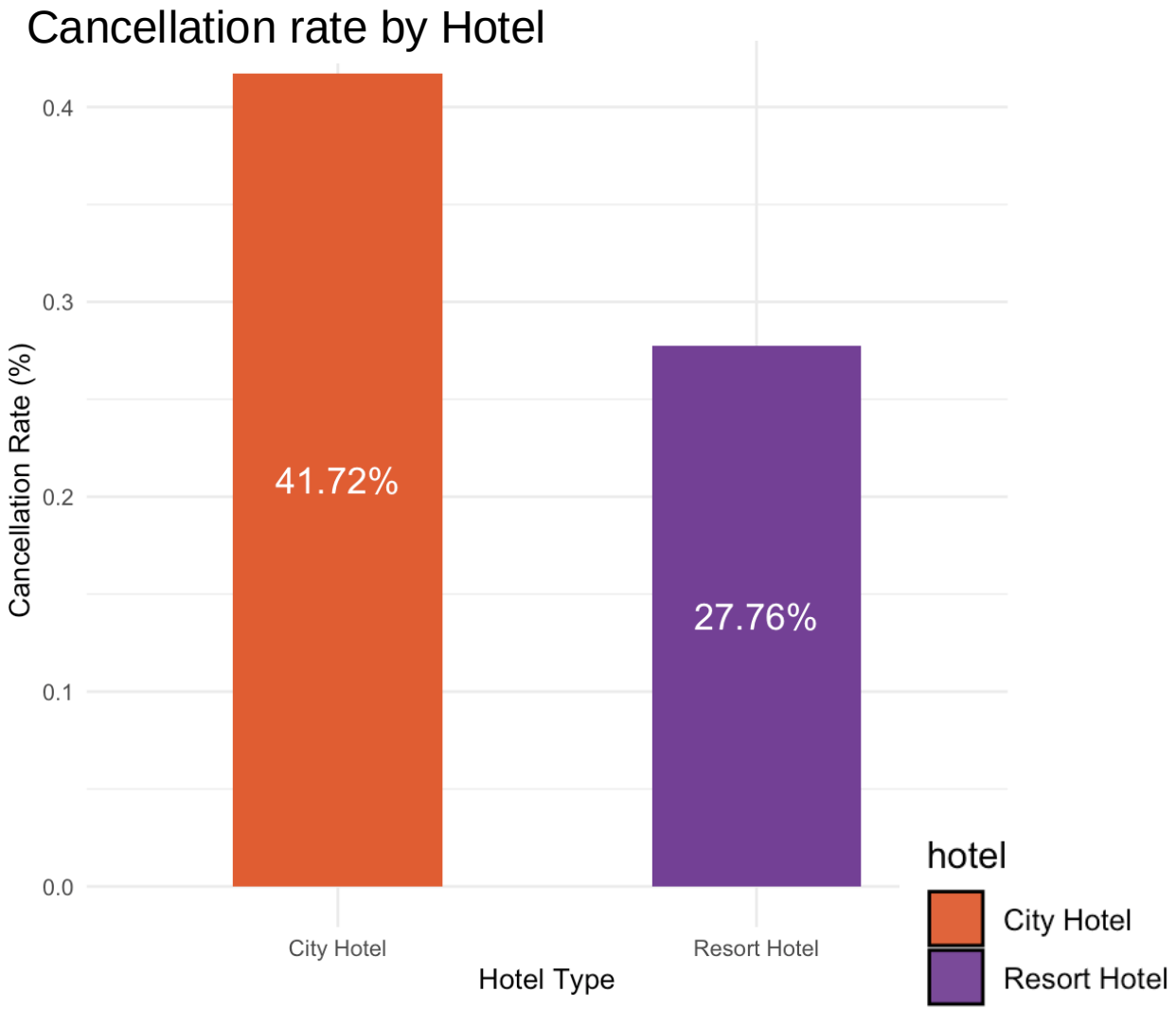


Chi-squared Test

X-squared = 0.04387
df = 11
p-value = 1

Cancellation rate by month range from 0.3 to 0.5, **No Significant Difference**

EDA - Cancellation Rate by Hotel

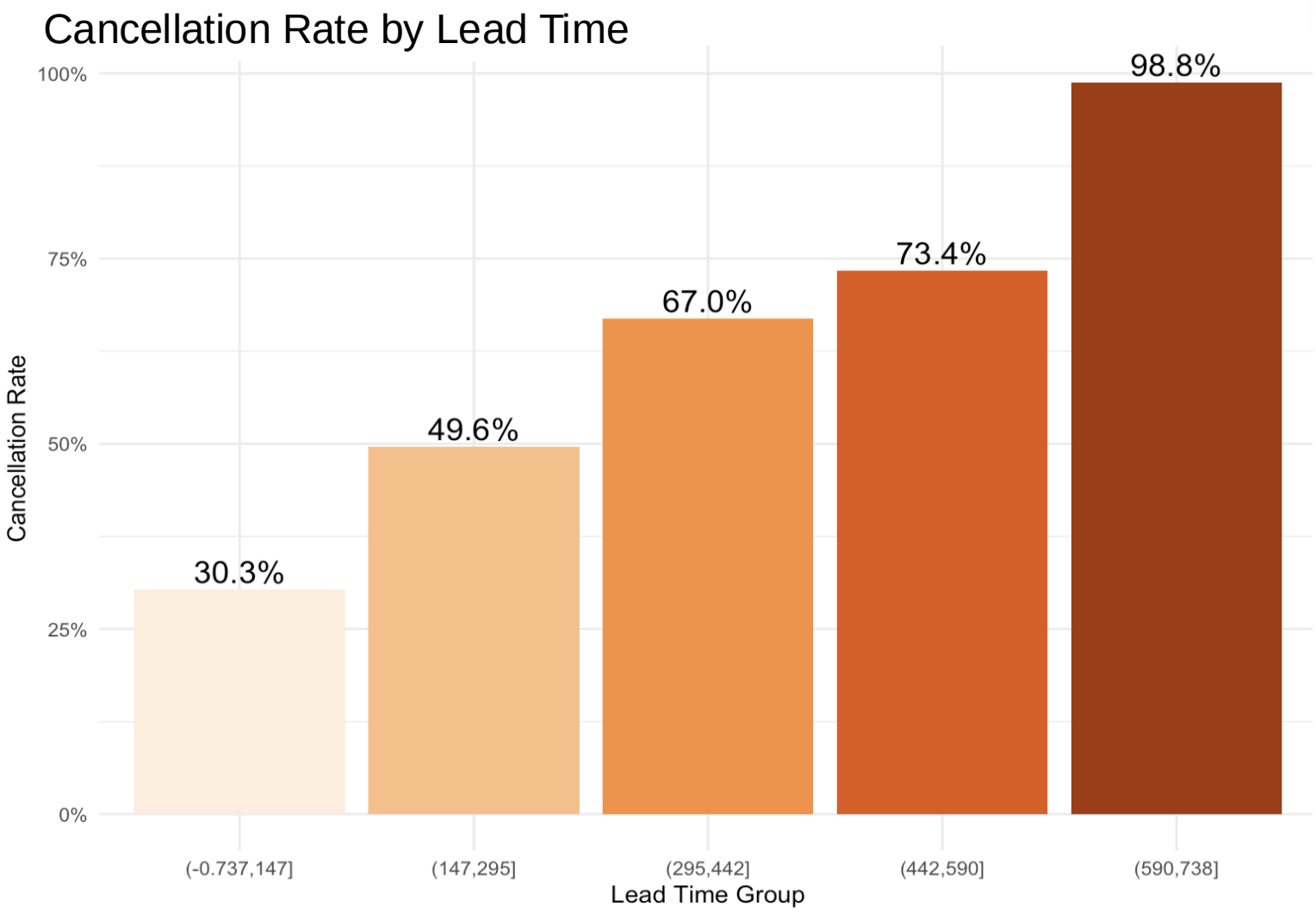


City Hotel has higher cancellation rate than Resort Hotel

Why?

Let's continue analysis…

EDA - Lead Time Analysis



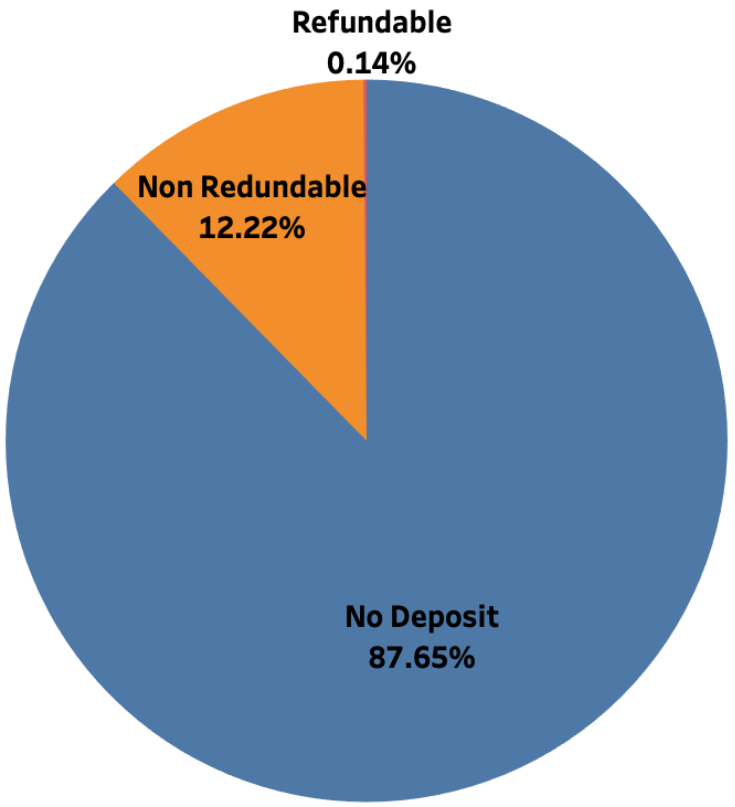
Lead Time Increases
Cancellation Rate Increases

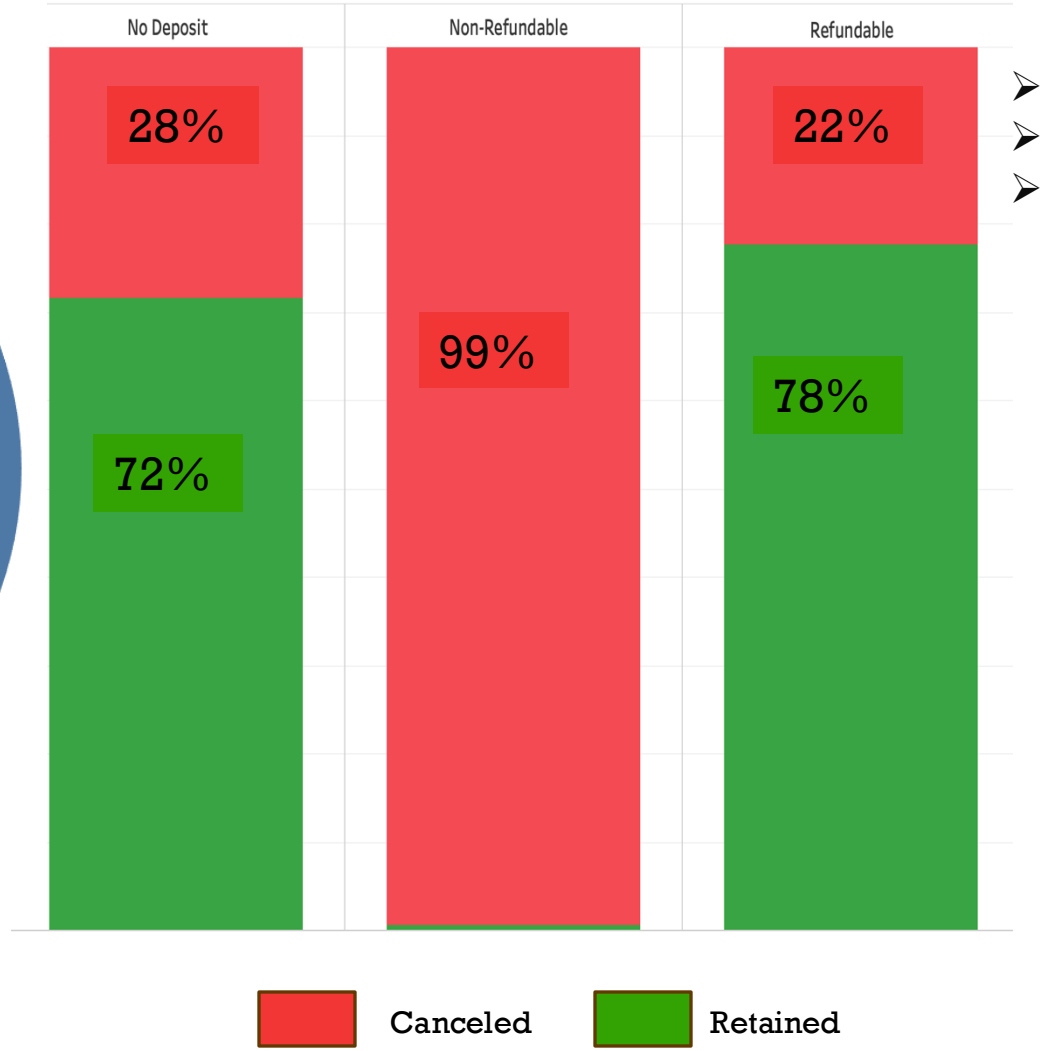
Potential Reasons:

- **Schedule Conflicts:** Customers booking far in advance are more likely to encounter changes in their schedules.
- **Financial Changes:** Long lead times might also see shifts in customers' financial situations, affecting their ability to commit to the booking.

Lead Time: Number of days that elapsed between booking and planned arriving date

EDA - Deposit Type Analysis

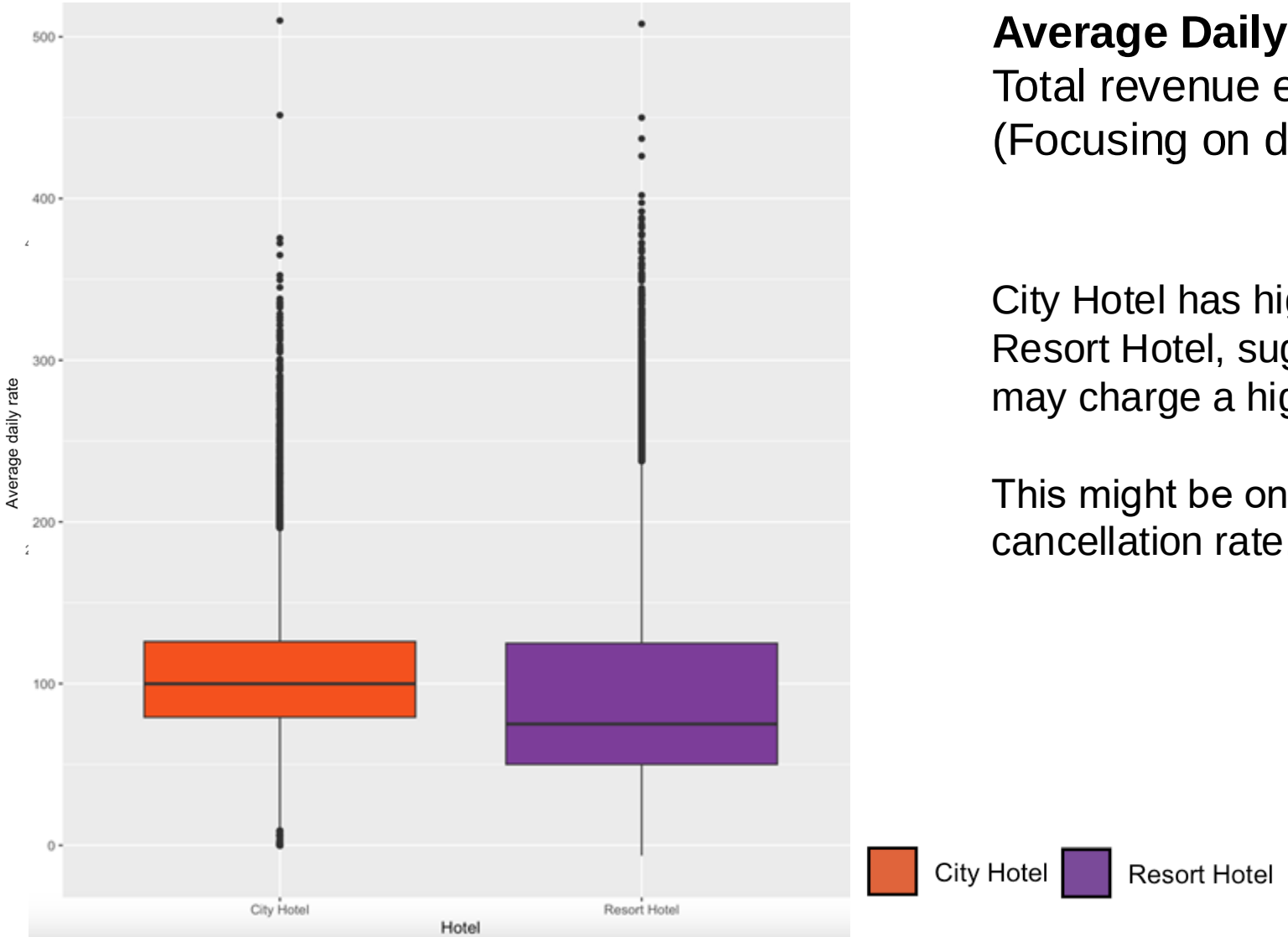




- Most customers do not put any deposits
- Non- refundable: high cancelation
- Refundable and No Deposit: low cancelation, indicates loyalty

EDA - Average Daily Rate Analysis

Average Daily Rate by Hotel

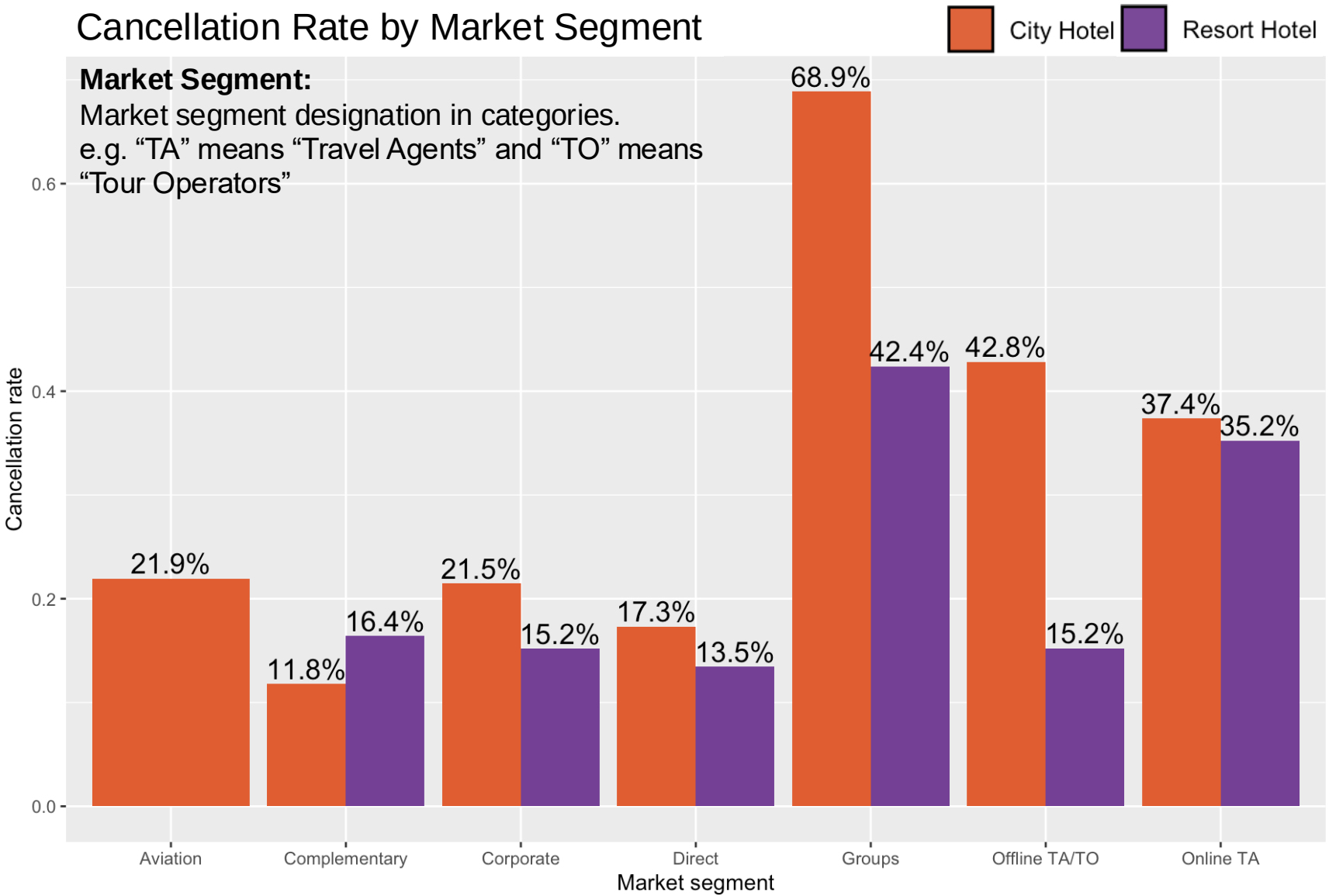


Average Daily Rate (ADR) :
Total revenue earned /Total number of staying nights (Focusing on daily prices)

City Hotel has higher median ADR than that for the Resort Hotel, suggesting that, on average, the City Hotel may charge a higher daily price.

This might be one of the reasons for City Hotel's high cancellation rate.

EDA - Market Segment Analysis



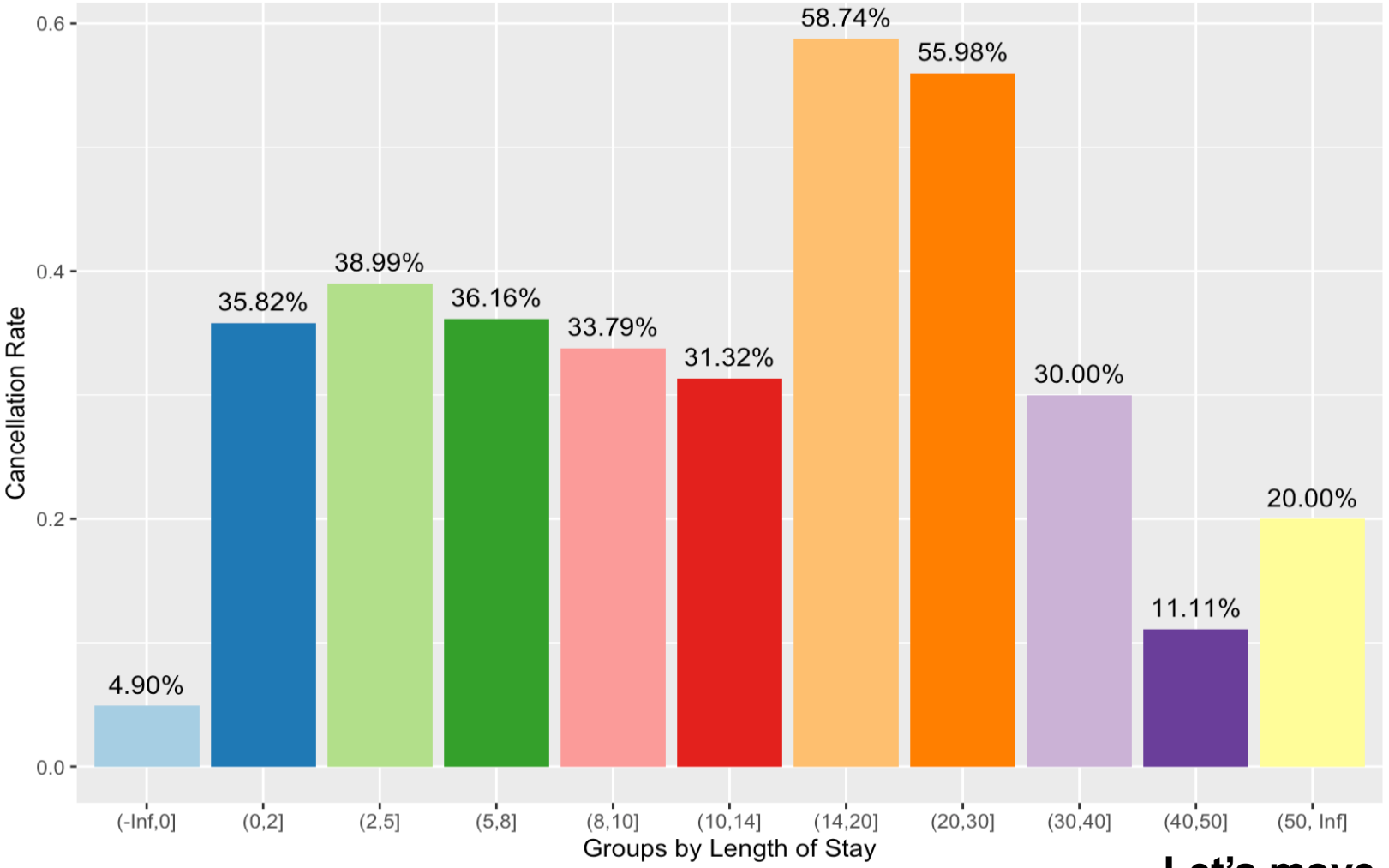
No matter City or Resort, 'Groups' has the highest cancellation rate

'Complementary' and 'Corporate' have lowest ones

This suggests that group bookings are most unstable, while corporate and complimentary offers are more reliable

EDA - Length of Stay Analysis

Cancellation Rates by Planned Length of Stay



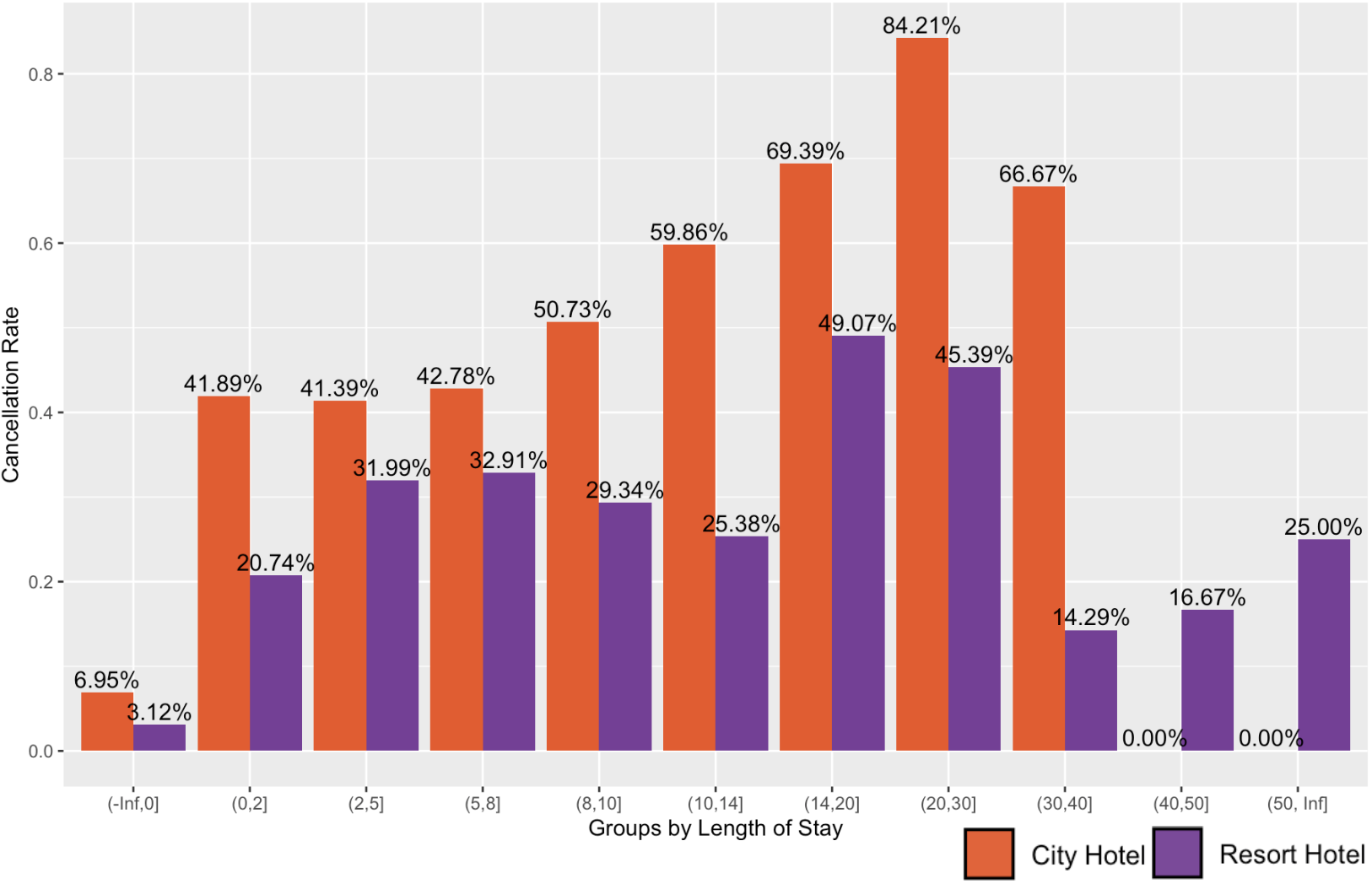
Short-term and very long-term bookings show lower cancellation rates.

There might be an optimal booking length from a risk management perspective.

Let's move on to see the difference in two hotels…

EDA - Length of Stay Analysis by Hotel

Cancellation Rates by Planned Length of Stay by Hotel

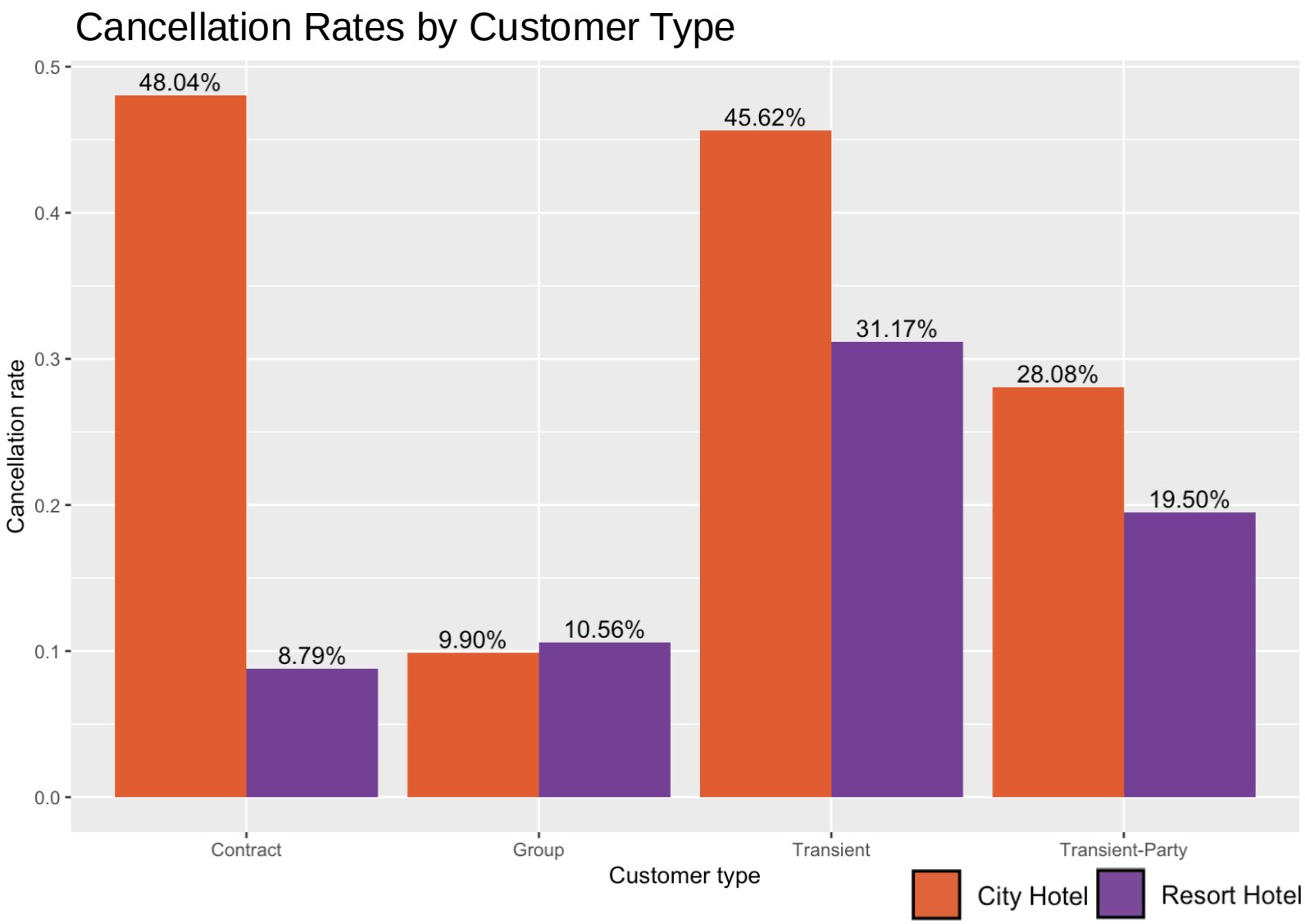


City Hotel has higher cancellation rate for each stage, and nearly zero travelers for booking longer than 40 days according to the dataset

Implications for Hotel Strategy:
Potential issues with long-stay bookings at City Hotels could indicate a need for improved management practices or customer engagement strategies.

Resort Hotels may benefit from promoting longer stays as their cancellation rates are lower for extended periods.

EDA - Customer Type Analysis

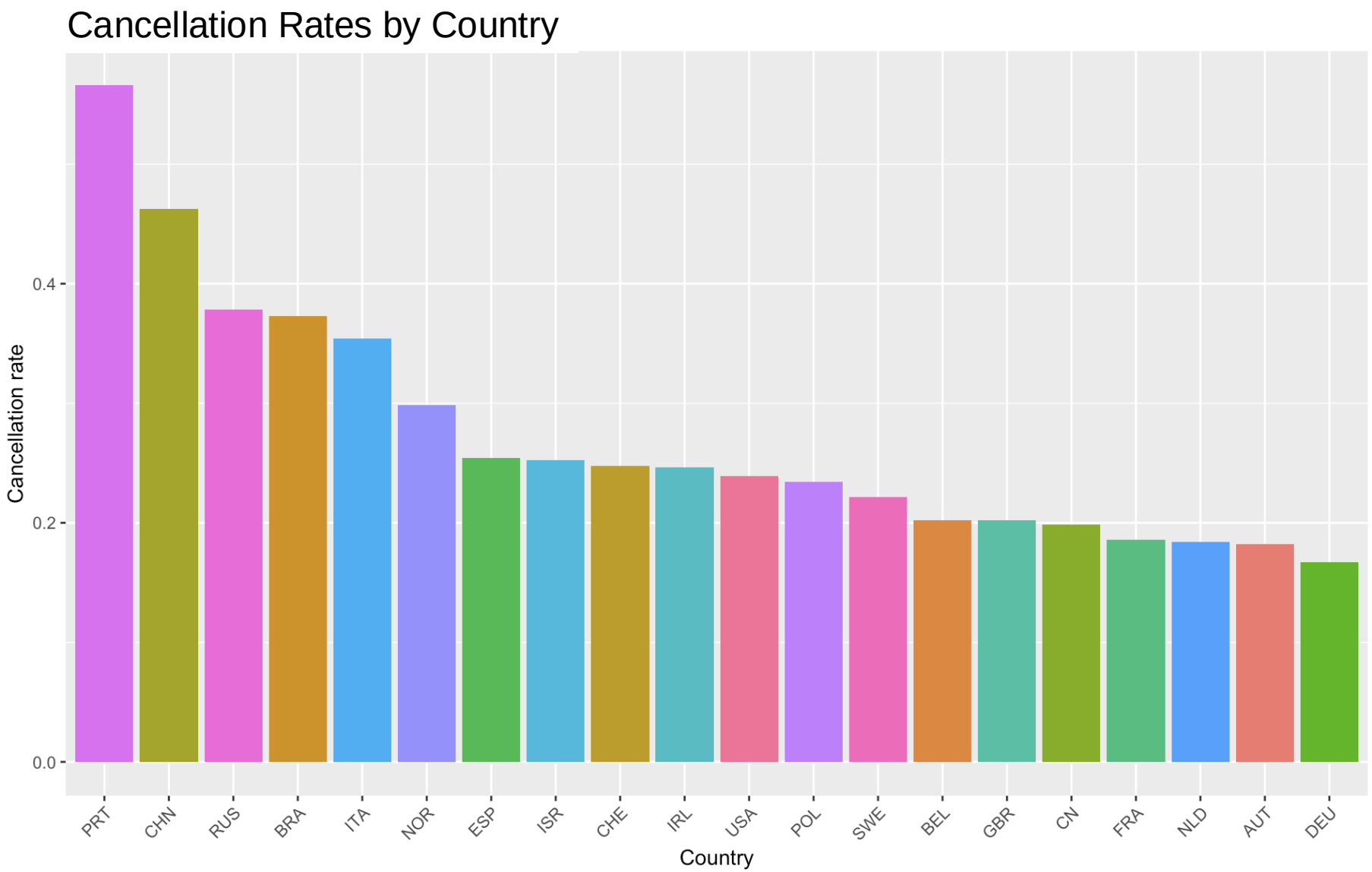


City Hotel has higher cancellation rate for 'Transient', 'Transient-Party' and 'Contract'

Resort Hotel has higher cancellation rate for 'Group'

This suggests that two hotels' clients verify and different type of customers have different behavior patterns.

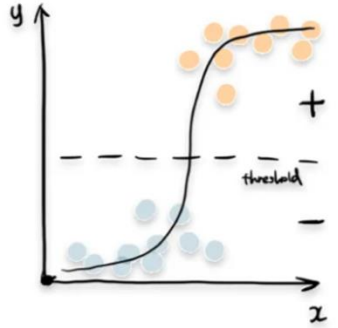
EDA - Country Analysis



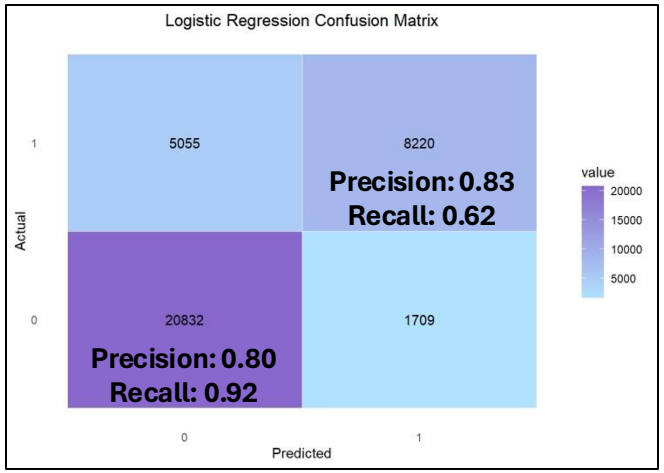
Portugal (PRT), where the hotel is located, has the highest cancellation rate. This could suggest that local customers are more likely to cancel their bookings, perhaps because of the convenience of proximity.

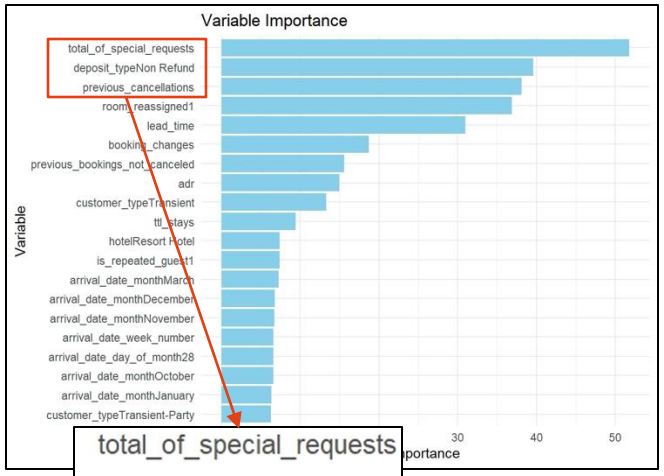
Cancellation rates verifying by country, which might indicate a variety of factors including the cost of travel, the purpose of visit, or the reliability of plans.

Logistic Regression Model

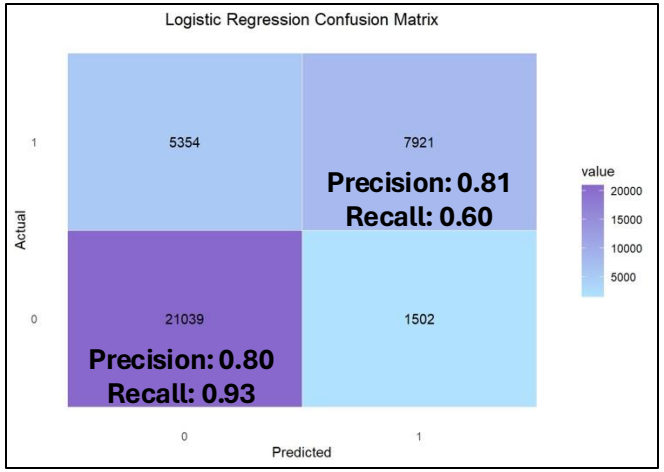


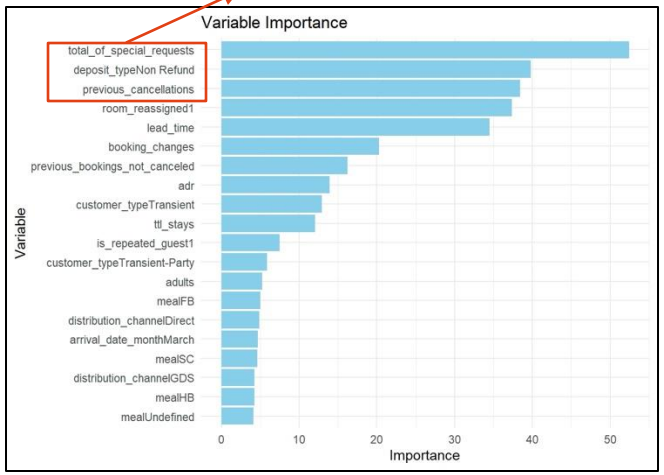
All variables – Accuracy: 0.81





Reduced variables – Accuracy: 0.81

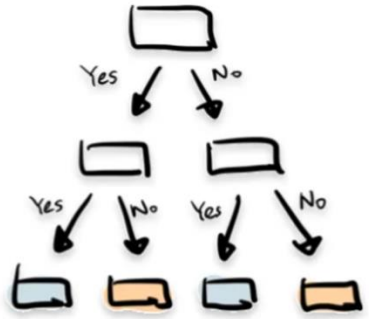




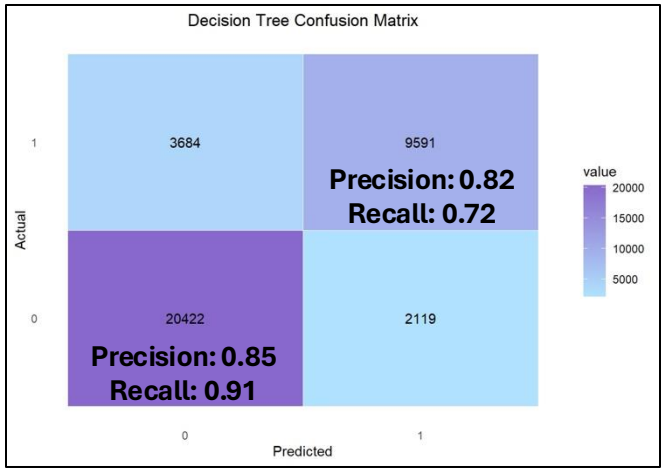
total_of_special_requests
deposit_typeNon Refund
previous_cancellations

- ✓ Both datasets (all and reduced variables) can reach 0.81 accuracy.
- ✓ ‘total_special_request,’ ‘deposit_type,’ and ‘previous_cancellation’ are the key variables in logistic regression model

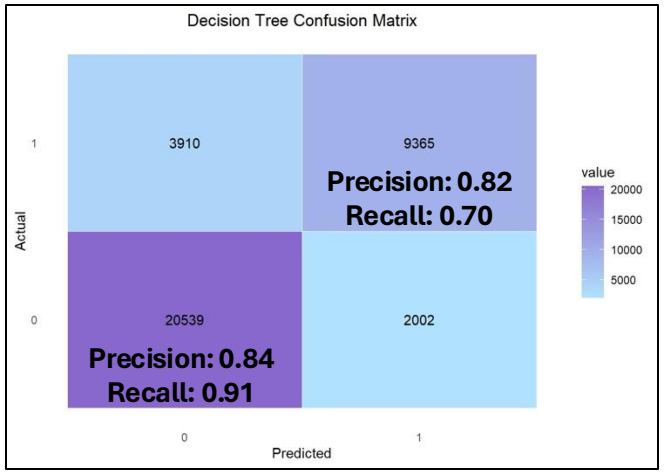
Decision Tree Model

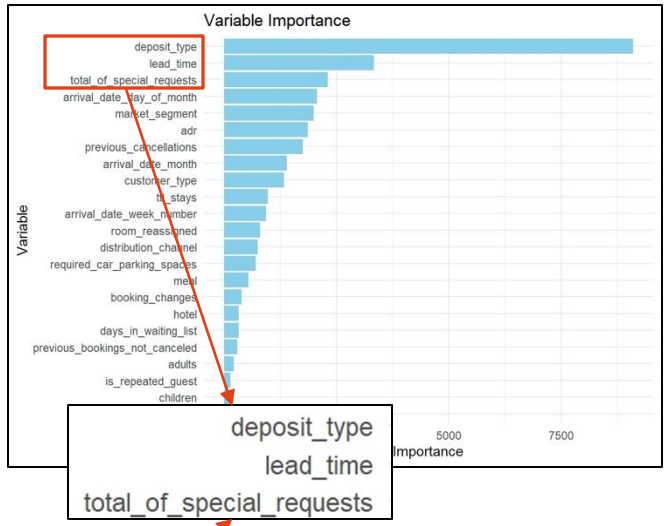


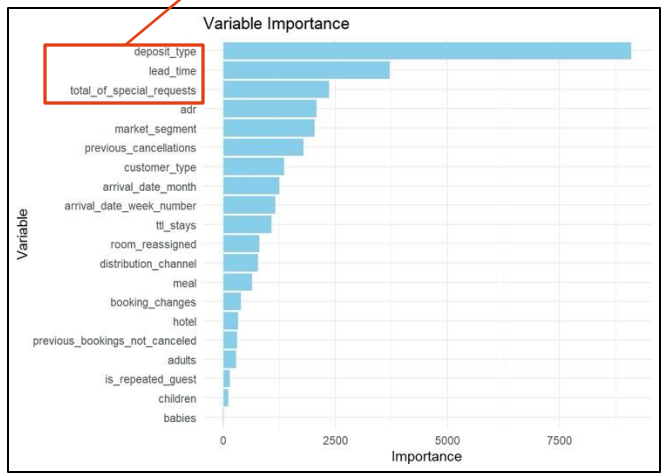
All variables – Accuracy: 0.84



Reduced variables – Accuracy: 0.84



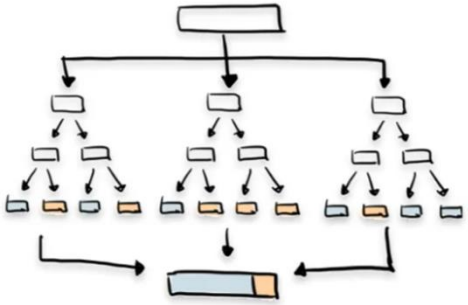




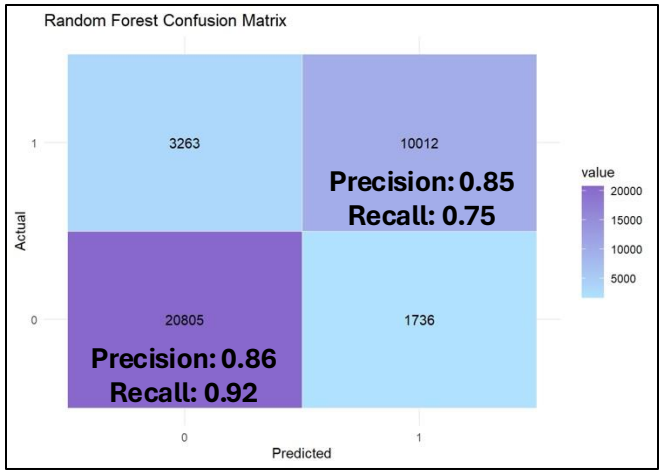
deposit_type
lead_time
total_of_special_requests

- ✓ Accuracy is slightly better than the Logistic Regression model.
- ✓ ‘Canceled’ & ‘Non-canceled’ can both have good precision and recall in the Decision Tree Model.
- ✓ ‘deposit_type,’ and ‘lead_time’ are the key variables in the Decision Tree model.

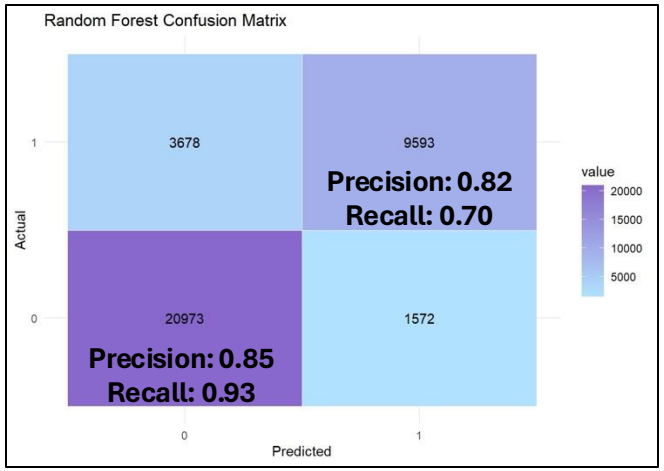
Random Forest Model

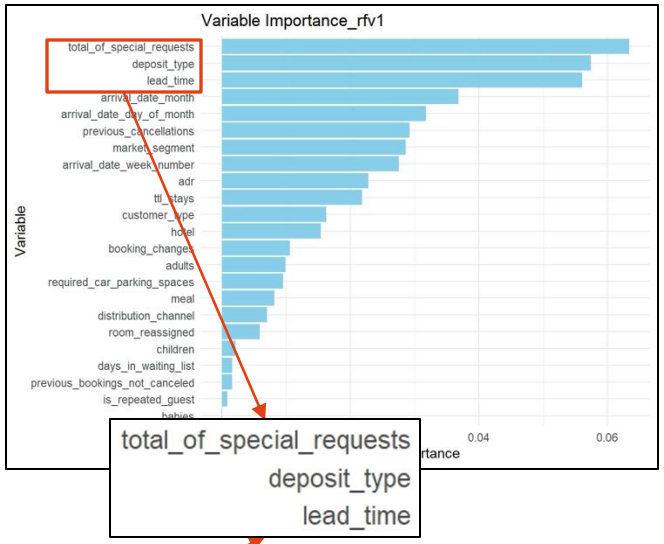


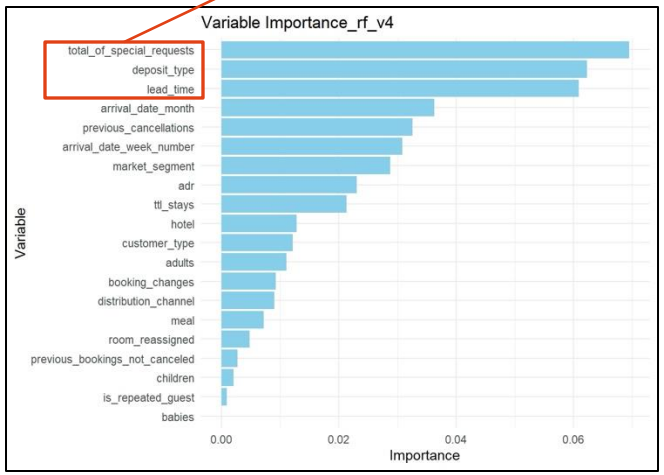
All variables – Accuracy: 0.86



Reduced variables – Accuracy: 0.85







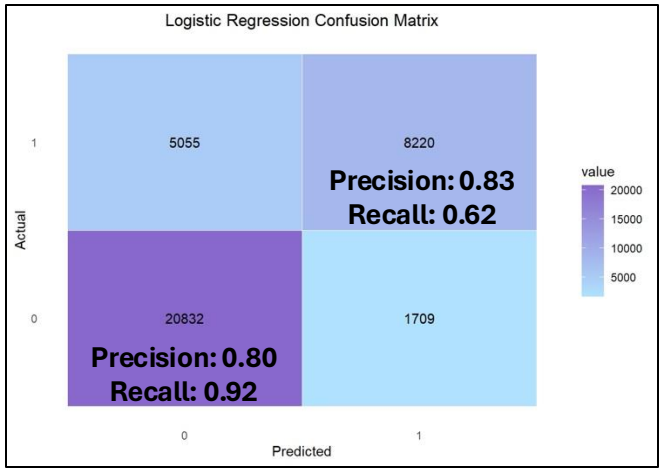
total_of_special_requests
deposit_type
lead_time

- ✓ Accuracy is slightly better than the Logistic Regression model.
- ✓ Precision and Recall are slightly improved in both categories compared to the Decision Tree model.
- ✓ ‘total_special_request,’ ‘deposit_type,’ and ‘lead_time’ are the key variables in the Random Forest model.

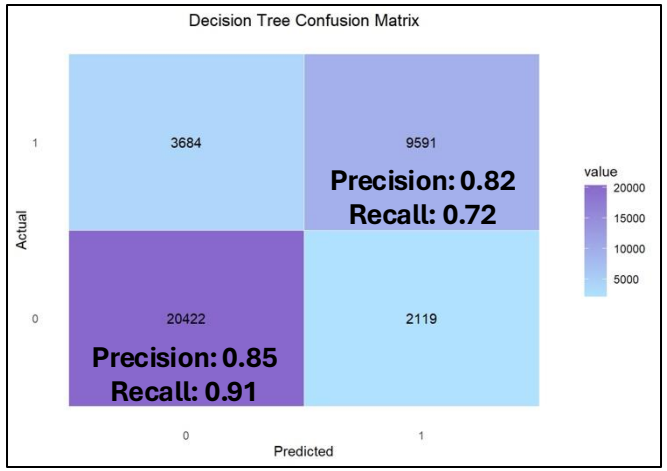
Model Comparison

- ✓ The Random Forest model demonstrates the highest accuracy among all models considered.
- ✓ In addition to accuracy, precision and recall metrics in both categories are also crucial. The Random Forest model exhibits higher precision and recall across both categories.
- ✓ Precision in the 'non-canceled' category is particularly significant for hotel owners. High precision in this category enables more accurate forecasting of customer arrivals, allowing for better preparation and management of resources.

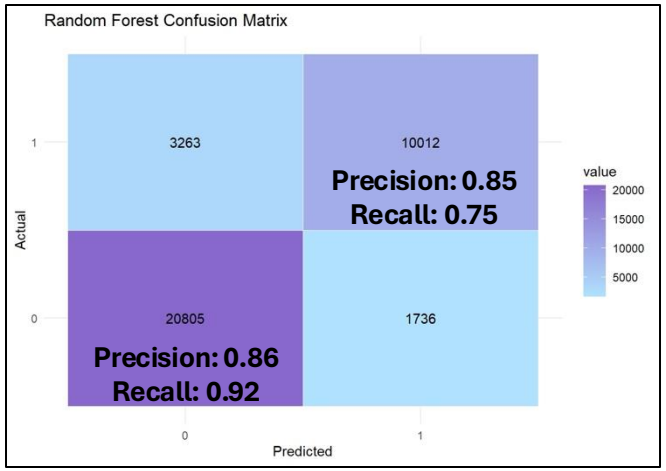
Logistic Regression Accuracy: 0.81



Decision Tree Accuracy: 0.84

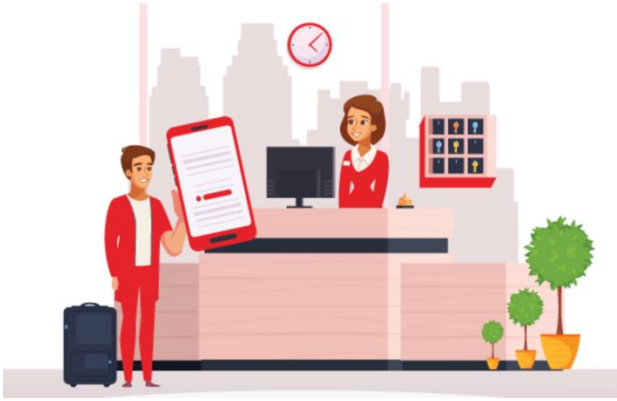


Random Forest Accuracy: 0.86



Recommendations

❖ **Promote "special request options" on booking platforms:**
- ✓ When request are met, less inclined to cancel.
- ✓ Foster customer retention for future booking.



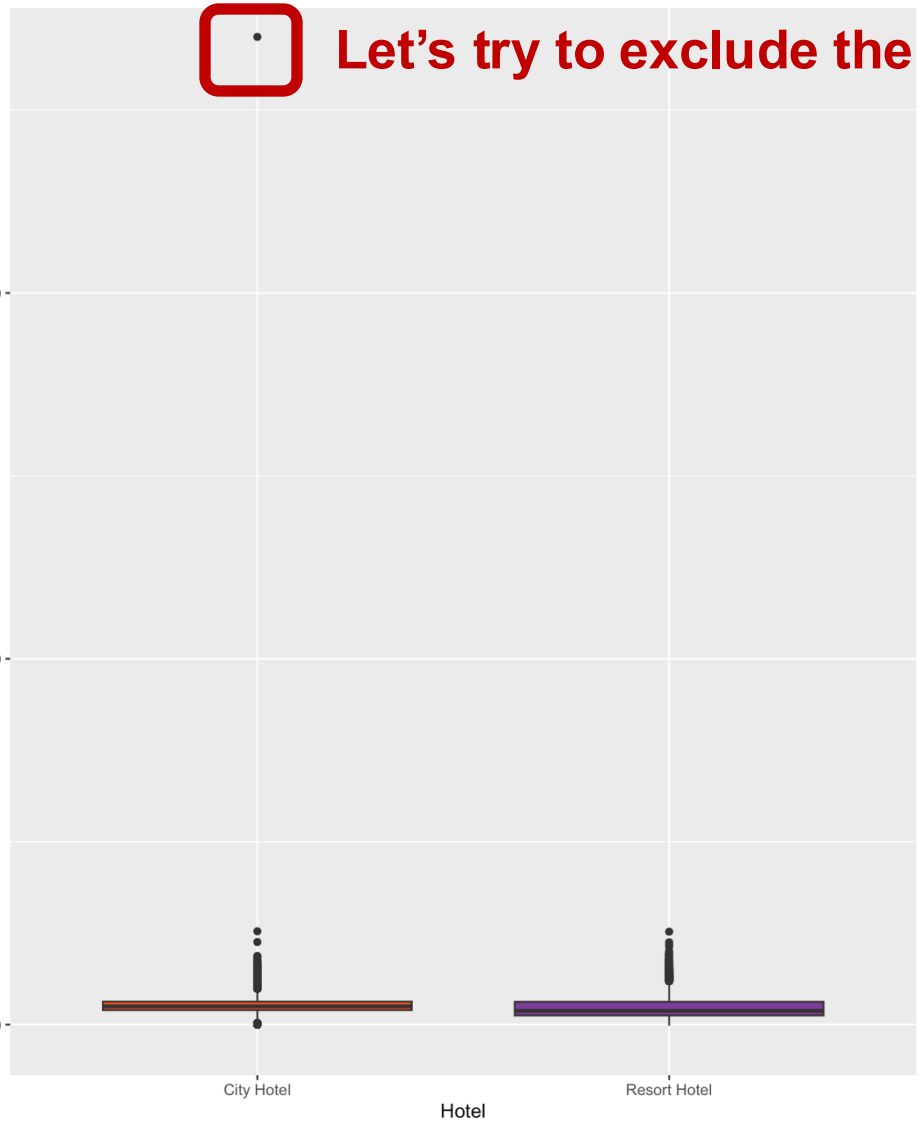
❖ **Flexible rescheduling option**
- ✓ Unusual trend: 99% non-refundable bookings cancelled
- ✓ Offer option to reschedule without penalty
- ✓ Help retain customers while preserving revenue

❖ **Maintain effective communications with guests:**
- ✓ Email/call for reminders ahead of bookings
- ✓ Manage cancellations and fill vacancies

❖ **Special offers for local tourists:**
- ✓ 60% of the local guests canceled bookings
- ✓ Personalized offers and loyalty programs to encourage repeat bookings

Appendix - Average Daily Rate Analysis

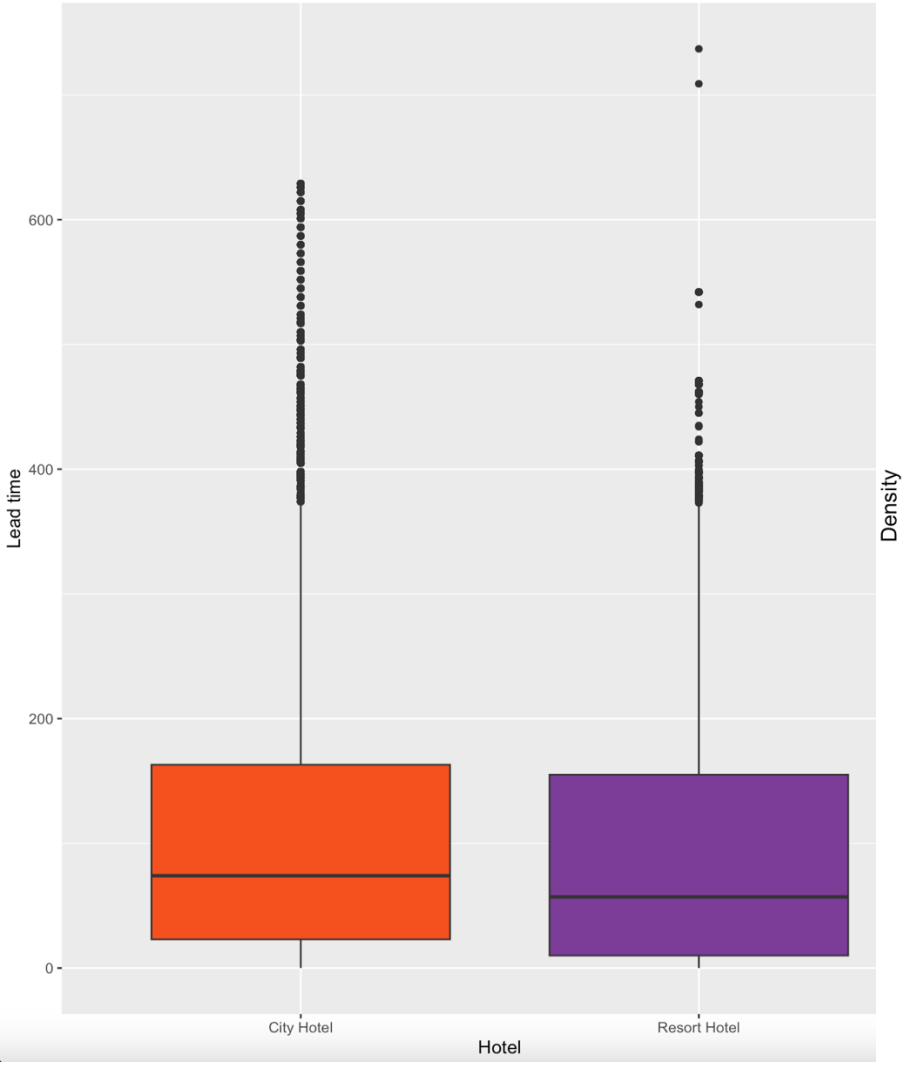
Average Daily Rate by Hotel



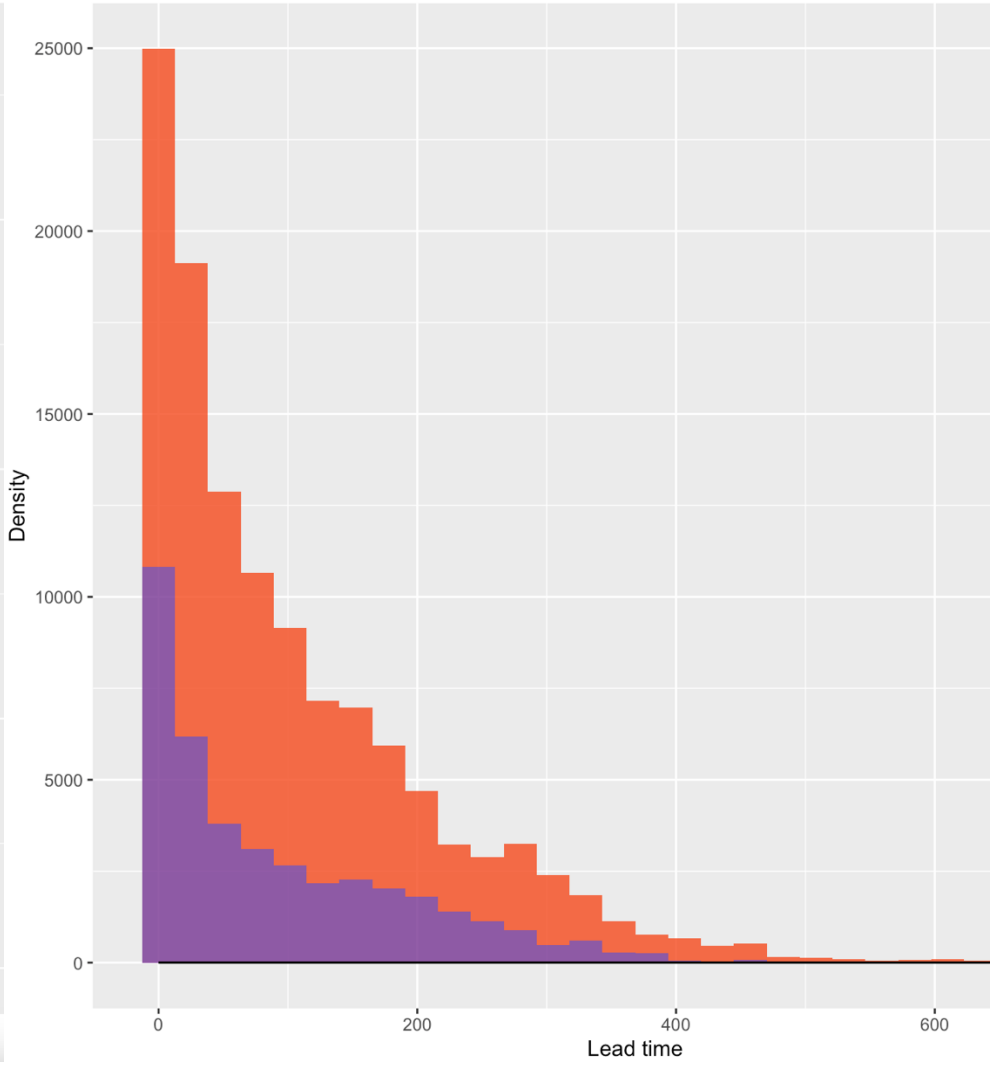
Average Daily Rate (ADR) :
Total revenue earned /Total number of staying nights

Appendix - Lead Time Analysis

Lead Time by Hotel



Lead Time Distribution by Hotel



City Hotels has longer lead times in bookings, which could correlate with higher cancellation rates as previously discussed.

Resort Hotels may attract more short-notice bookings, which might be associated with a more stable and predictable guest turnover.

Lead Time: Number of days that elapsed between booking and planned entering date