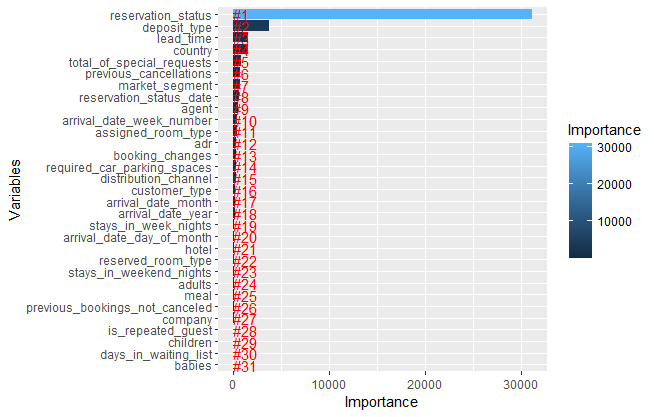
**Random Forest Model:**

The Random Forest model is popular because it’s powerful, flexible, and relatively easy to use. It can handle both regression and classification problems, and it works well with both small and large datasets. It also handles a mix of categorical and numerical features, and it’s robust to outliers and missing values

model <- randomForest(is\_canceled ~ ., data = train, ntree = 50)



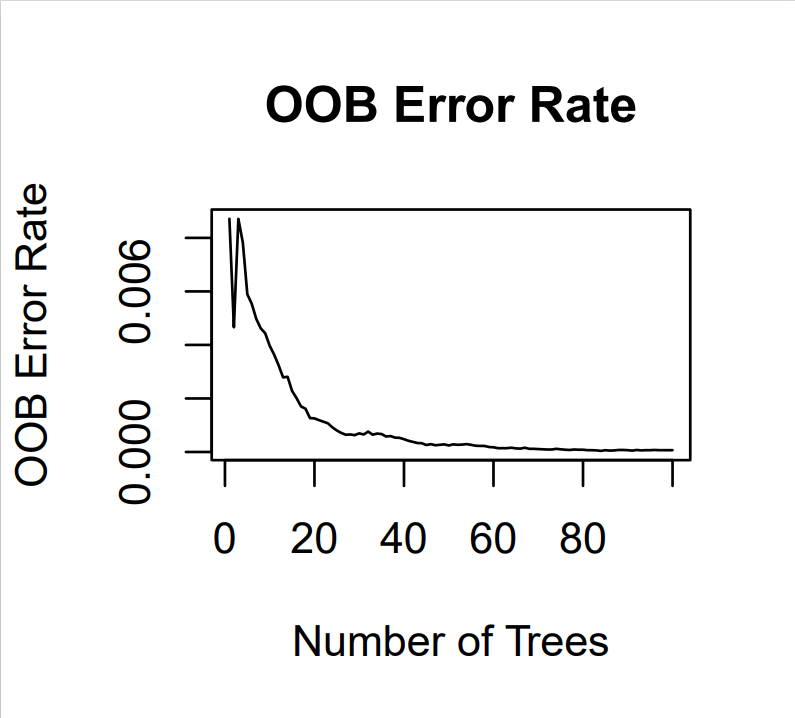
This graph is a visualization of the importance of each variable in your Random Forest model. The y-axis represents the variables used in the model, and the x-axis represents the importance of each variable.

The importance of a variable in a Random Forest model is measured by the increase in the model’s prediction error after permuting the variable. A higher increase in error indicates a more important variable.

From the graph, it appears that **reservation\_status, deposit\_type,** and **lead\_time** are the most important variables in predicting whether a guest will cancel their reservation. These variables have the highest importance scores, which suggests that they have the strongest impact on the model’s predictions.

On the other hand, variables like babies, children, waiting\_list, and guests have lower importance scores, which suggests that they have less impact on the model’s predictions.

This information can be very useful in understanding which factors are most influential in a guest’s decision to cancel their reservation. For example, if reservation\_status is a significant predictor, the hotel might want to investigate their reservation system and see if there are any aspects that could be improved to reduce cancellations. Similarly, understanding the impact of deposit\_type and lead\_time could help the hotel develop more effective policies and strategies to reduce cancellations.

****