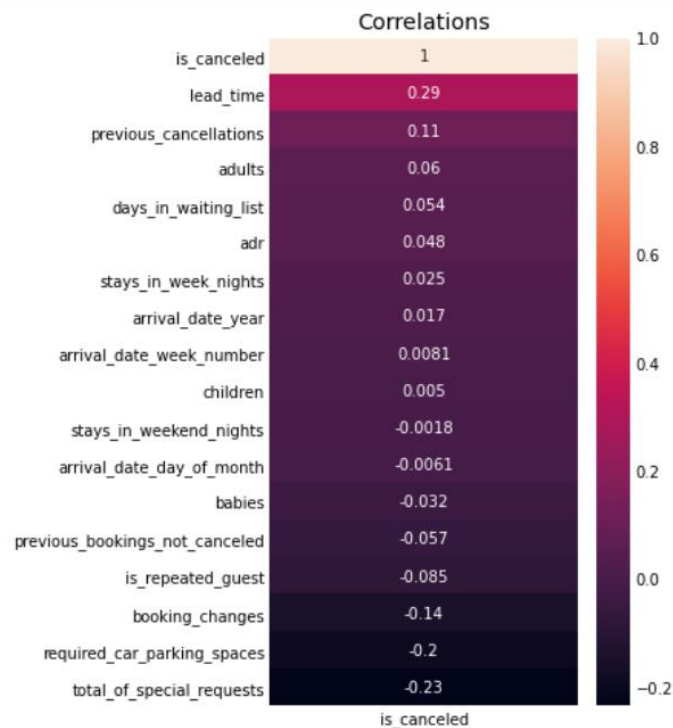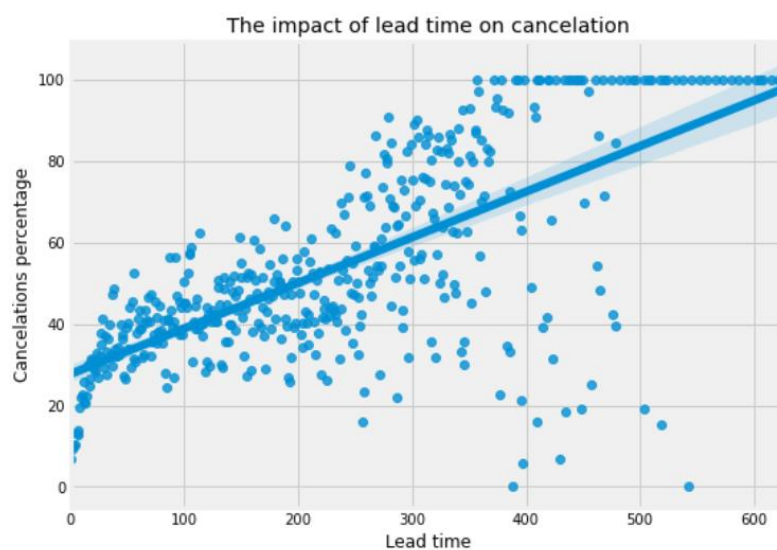The goal of this project is to predict whether a hotel booking will be cancelled or not. To achieve this goal, first we need to understand how other features are related to the target variable (is_canceled).
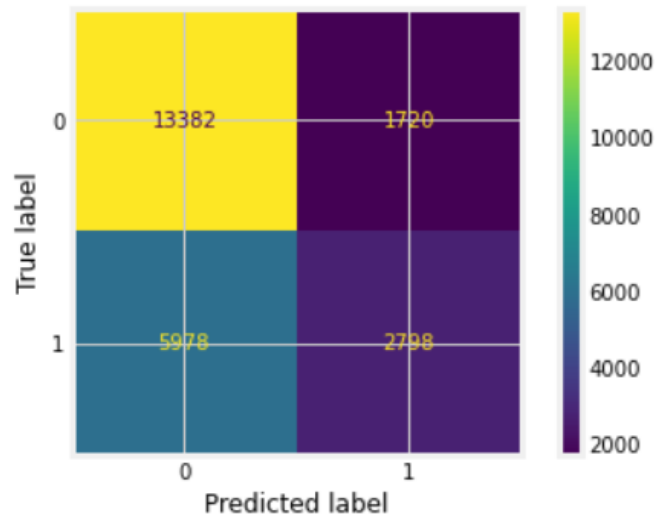

The heat map result suggests that lead time "the number of days between booking date is arrival date" has the highest positive impact on the reservation status.

Insight: When customers have more time to cancel, the likelihood if cancellation increases.

I construct a logistic regression model based on lead time feature.
The evaluation of the model is illustrated in the following confusion matrix plot.

The result is a bit disappointing; it seems that using logistic regression with only that one feature was not the best idea.

So, in the project I will consider other Machine learning algorithms. I think Decision Tree might be the one!