Predicting Hotel Booking Cancellations

By: Jazmaine Vitta

Improving Hotels Operations and Revenues

The Problem

Can the likelihood of **Resort and City Hotel booking cancellations** be predicted to help companies be adequately equipped to manage customers and aid in **reducing the businesses annual operational costs** through overhead rationalization and operational improvements?

The Negative Effects of Booking Cancellations

- Causes challenges for accurately forecasting the hotels daily demand

- Rooms that have been cancelled or “No Showed” are often resold for a lesser rate or worse not resold at all

A booking cancellation may not seem like a problem to the naive. That is until you factor in the cost of staff and other necessities.

The Answer

To combat the negative effect of cancellations, hotels need to be able to accurately predict the likelihood that a booking will be canceled or not.

By using my classification model I can assist hotels to overcome the uncertainty that comes with booking cancellations and increase revenue for the business.

The Data

From: Hotel Booking Demand

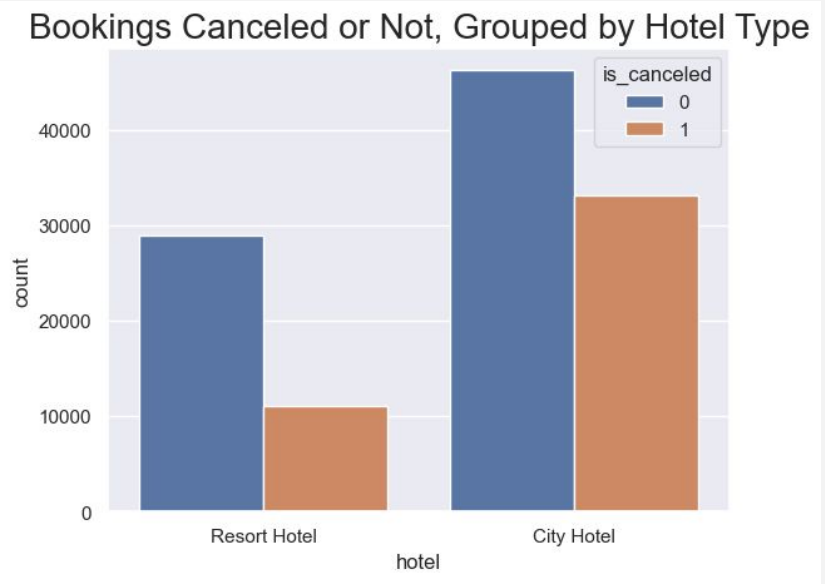
by Jesse Mostipak on Kaggle.com

The data contains City Hotel & Resort Hotel bookings due to arrive between the years of 2015 - 2017.

The data is made up of a total of 119,389 booking observations with 32 columns

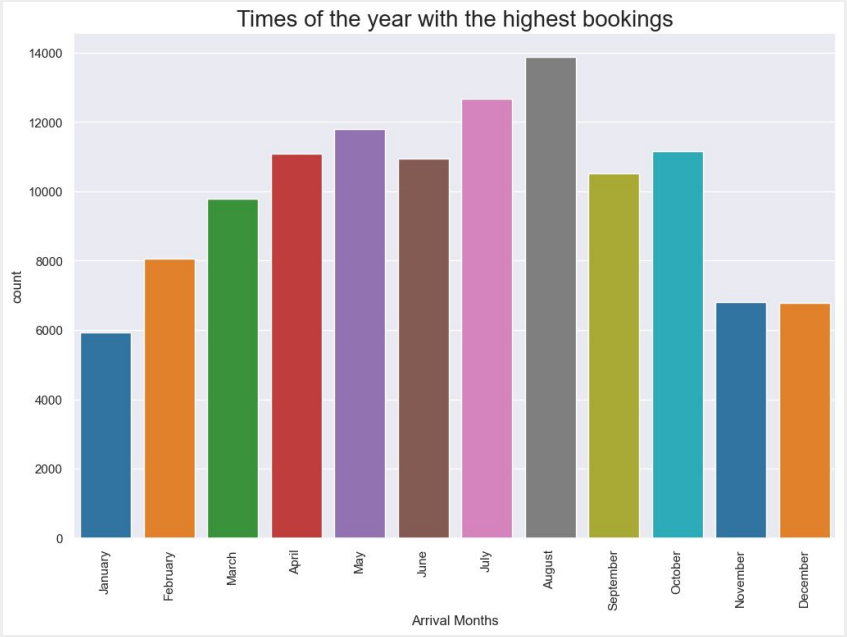
What The Data Shows

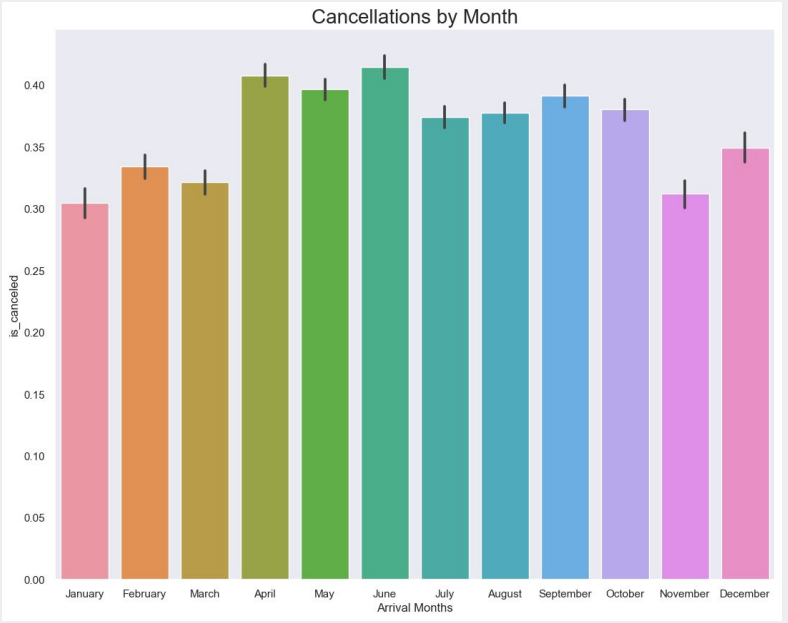
- The data contains almost **double** the amount of City Hotel bookings compared to Resort Hotel bookings
- City hotels experience almost as many cancellations as they do bookings, which is a big contrast to resort hotels who receive around a third of their bookings canceled.



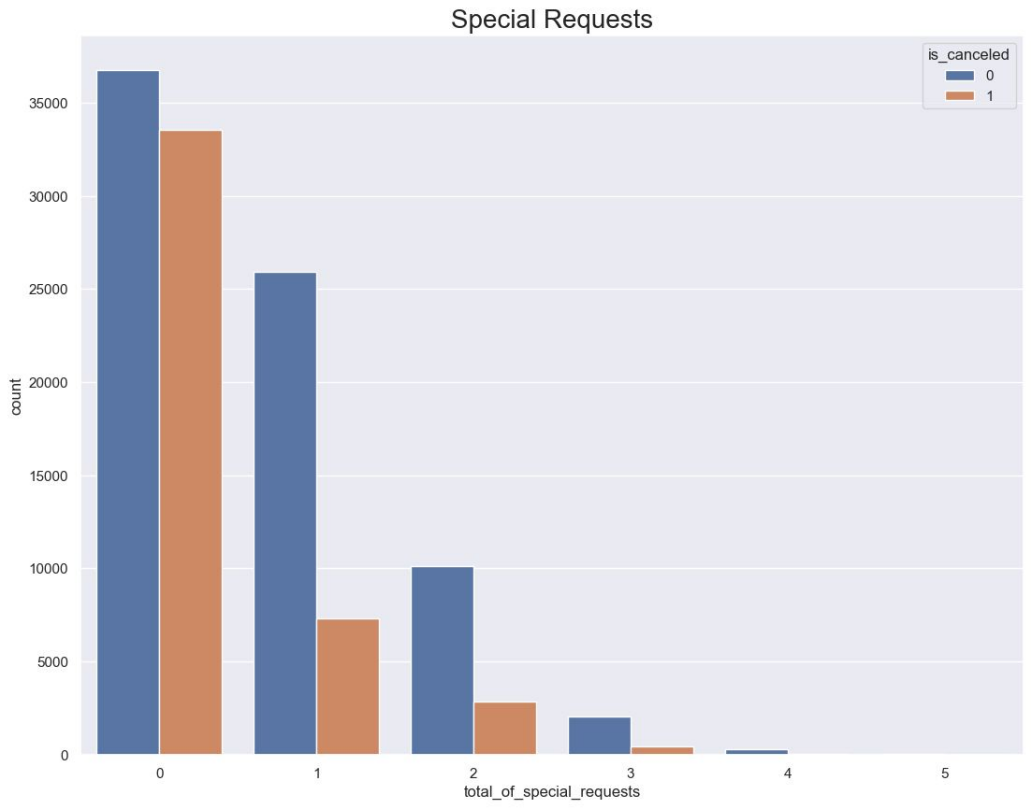
What The Data Shows

- Although the summer months are the time of the year with the most bookings, the spring is when hotels experience the most cancellations



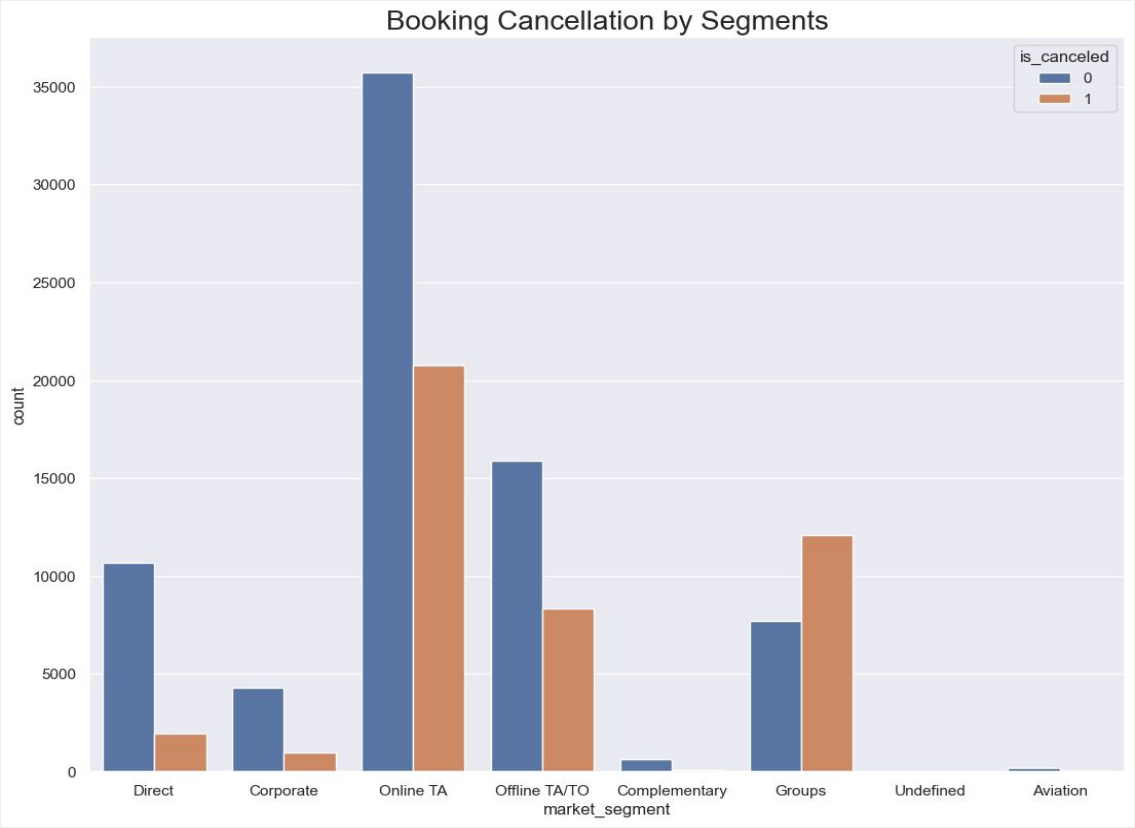


What The Data Shows



Customers with bookings that have one or more special requests are much less likely to cancel their trip!

What The Data Shows



Based on where reservations are made: **Groups** are more than less likely to cancel. **Direct** bookings (made directly with Hotel) are much less likely to cancel

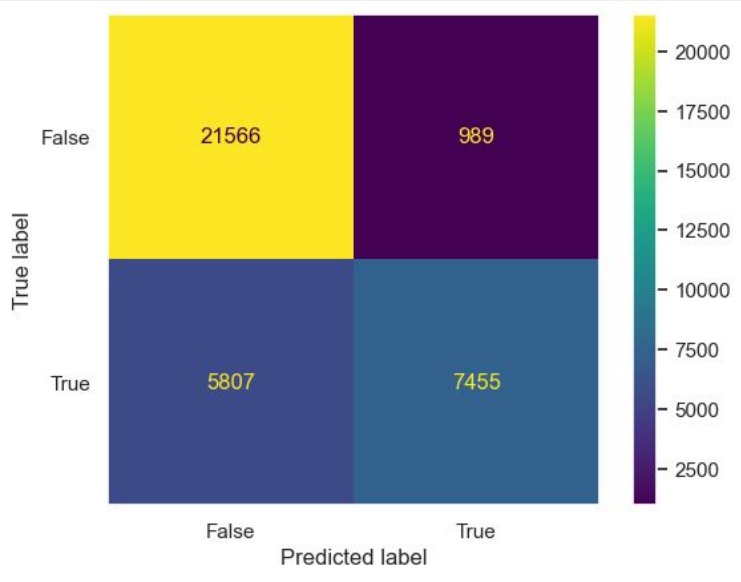
What The Data Shows

As shown in the heatmap below, the features most correlated to '***is_canceled***' which is the target feature are '***lead_time***', '***total_of_special_requests***', '***booking_changes***' and '***required_car_parking_spaces***'.

Model Selection

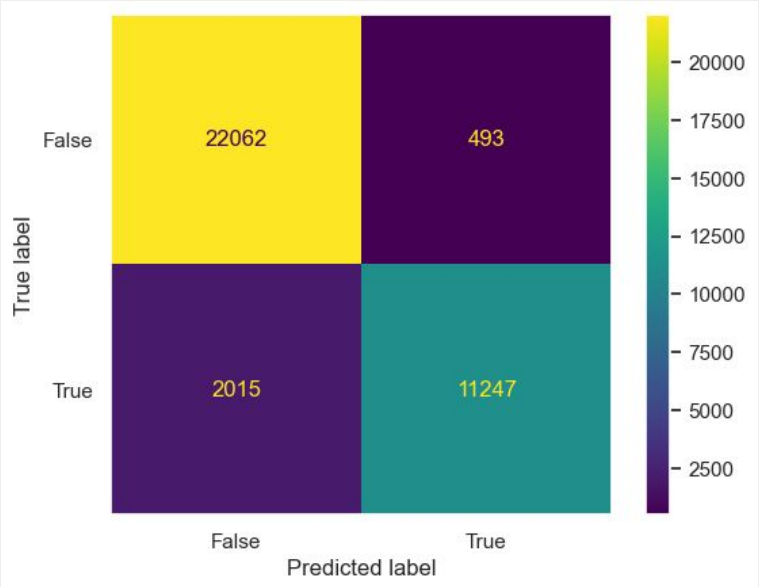
Logistic Regression Confusion Matrix

Accuracy Score of Logistic Regression is : 0.8102576988580842



Random Forest Confusion Matrix

Accuracy Score of Random Forest is : 0.9299773850406232



Why Random Forest Classifier / Model Selection

The random forest algorithm implements both Classification and Regression problems in ML. The Random Forest is a classifier that includes several decision trees instead of relying on one decision tree, the random forest takes the prediction from each tree, and based on the majority votes of predictions, it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

Summary & Conclusions

- Predicting cancellations is a real problem for the hotel industry!
- Having a good understanding of this problem and the features that closely relate with cancellations will be very useful to your business.
- The model I created will no doubt assist in decreasing the possibility of your business losing revenue by being under/over prepared.
- My Random Forest Model has a high accuracy of 93.09%, to predict hotel booking cancellation.

References

Link to Kaggle Dataset