

## Capstone Project 1 Proposal

According to <http://blog.experience-hotel.com/>, the average rate of hotel cancellation from all sources is about 24%. Consequently, it has always been the hotel revenue manager's biggest concern to counteract hotel cancellation. Building an appropriate model that predicts the likelihood of hotel cancellation helps revenue managers to better understand the situations and propose policies to accommodate or avoid hotel cancellation.

The main data that is made available for me is from Kaggle (<https://www.kaggle.com/jessemostipak/hotel-booking-demand>). The data set contains data from 2 hotels located at Lisbon, Portugal: city hotel and resort hotel. The data set specifies many features such as the cancellation status, time stayed, check in/out date, etc. Along with the first data set, another data set will be added to the model in a suspicious manner having a huge effect on hotel cancellation; the added data is obtained from Lisbon, Portugal weather records. (<https://www.timeanddate.com/weather/portugal/lisbon/historic?month=7&year=2015>). The data will be scraped directly from the website. All the works will be posted on github. The following flow chart shows the steps to solve the problem:

