数据准备，我对数据做了一些EDA和缺失值检查，选取了全部变量作为feature。把数据分为测试集和训练集，其中测试集比例为20%。 建模，选择了logistic regression, XGBoost, Random Forest, Extra Trees, AdaBoost, GBDT, SVM，利用10折交叉检验，选择最优参数，并给出测试集的AUC作为模型评估指标。大部分的模型的AUC都在0.78左右。通过观察模型的feature importance，我们可以发现，Price在模型预测中的影响较大，平均影响程度占到40%左右，Weekday，#Beds,Review也占据了10%以上的平均模型影响。我们评估了模型的AUC和训练速度，其中SVM由于训练速度过慢，Extra Trees，logistic regression的False Positive太高,并且AUC较低，因此不考虑加入下一层。 第二层模型，我们利用XGBoost，AdaBoost，GBDT输出概率值作为新的特征，并以新的特征值作为输入变量，利用XGBoost进行整合，希望提高模型稳定性，同样以测试集AUC为指标进行调参。 得到所有模型的参数后，我们以所有的数据为模型训练集以提高数据量，并利用训练出的模型对测试集做出预测。 总结：第一层：AdaBoost, XGBoost, GBDT（输出概率）；第二层：XGBoost(输出概率)。